PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

#14

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/15, 15/62, 15/11, 1/21, 5/10, 1/19, C07K 14/81, A61K 38/57, C07K 19/00, 16/38, A61K 39/395, A01K 67/027, C12N 15/00, G01N 33/68, C12Q 1/68 , A61K 31/70, C12Q 1/37

A3

(11) International Publication Number: WO 97/14797

(43) International Publication Date: 24 April 1997 (24.04.97)

(21) International Application Number: PCT/US96/16782

(22) International Filing Date: 18 October 1996 (18.10.96)

(30) Priority Data:
08/546,000 20 October 1995 (20.10.95) US

(71) Applicant: DANA-FARBER CANCER INSTITUTE [US/US]; 44 Binney Street, Boston, MA 02115 (US).

(72) Inventors: SOTIROPOULOU, Georgia; 89 Trenton Street, Boston, MA 02128 (US). ANISOWICZ, Anthony; 50 Upham Street, West Newton, MA 02165 (US). SAGER, Ruth; 20 Codman Road, Brookline, MA 02146 (US).

(74) Agents: DECONTI, Giulio, A., Jr. et al.; Lahive & Cockfield, 60 State Street, Boston, MA 02109 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(88) Date of publication of the international search report:
2 October 1997 (02.10.97)

(54) Title: CYSTATIN M, A NOVEL CYSTEINE PROTEINASE INHIBITOR

(57) Abstract

Isolated nucleic acid molecules encoding a novel cysteine proteinase inhibitor, cystatin M, which is a member of the Family 2 cystatins, are disclosed. Cystatin M exhibits cysteine proteinase inhibitory activity against papain, is downregulated in metastatic mammary epithelial tumor cells, as well as other tumor cells, and is upregulated in senescent cells. In addition to isolated nucleic acid molecules, the invention provides antisense nucleic acid molecules, recombinant expression vectors containing a nucelic acid molecule of the invention, host cells into which the expression vectors have been introduced and non-human transgenic animals in which a cystatin M gene has been introduced or disrupted. The invention further provides isolated cystatin M proteins, fusion proteins, antigenic peptides and anti-cystatin M antibodies. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided.

* (Referred to in PCT Gazette No. 46/1997, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

CYSTATIN M, A NOVEL CYSTEINE PROTEINASE INHIBITOR

Background of the Invention

Metastasis of a primary tumor is a multistage process involving numerous aberrant functions of the tumor cell. These aberrant functions include tumor angiogenesis, attachment, adhesion to the vascular basement membrane, local proteolysis, degradation of extracellular matrix components, migration through the vasculature, invasion of the basement membrane, and proliferation at secondary sites (Poste, G. and Fidler, I.J. (1980) *Nature* 283:139-146; Liotta, L.A. *et al.* (1991) *Cell* 64:327-336). Therefore, accumulative changes in the expression of multiple genes probably occur before tumor cells acquire the phenotype that enables them to metastasize. The identification of genes whose changes in expression determine the metastatic phenotype is essential for an understanding of the molecular mechanisms underlying metastasis and for the design of novel therapies designed to arrest progression of a primary tumor.

Increased proteolytic potential constitutes one well documented feature of the metastatic phenotype. This increased potential is thought to result from the combined aberrant regulation of proteolytic enzymes (*e.g.*, metalloproteinases, serine, cysteine and aspartyl proteinases) and their endogenous inhibitors (for a review, see *e.g.*, Sloane, B.F. and Honn, K.V. (1984) *Cancer Metastasis Rev.* 3:249-263). For example, the lysosomal cysteine proteinases cathepsins B and L are normally intracellular but when they are overexpressed in tumor cells, they may be secreted or become associated with the plasma membrane where they may act cooperatively in directly degrading components of the extracellular matrix and basement membrane and indirectly, by activating latent metalloproteinases. Alterations in intracellular trafficking and increases in expression and secretion of the lysosomal proteinases cathepsin B, D and L have been observed in a variety of malignant tumors, including breast carcinoma (Kolar, Z. *et al.* (1989) *Neoplasma* 36:185-189). Additionally, membrane-associated forms of cathepsin B and cathepsin L and of their endogenous low molecular weight Cysteine Proteinase Inhibitors (CPIs) may both play a role in the expression of the malignant phenotype. For example, increased activity of cathepsins B and L may result from reduced expression and/or activity of CPIs of cathepsin B and L (Sloane, B.F. *et al.* (1990) *Biol. Chem. Hoppe Seyler* 371:193-198; Lah, T.T. (1989) *Biochim. Biophys. Acta* 993:63-73; Steahan, K. *et al.* (1989) *Cancer Res.* 49:3809-3814).

Cystatins are the endogenous inhibitors of mammalian lysosomal cysteine proteinases, such as cathepsins B, L, H, and S, and the plant cysteine proteinases papain, actinidin, and ficin found both intracellularly and extracellularly (for reviews see *e.g.*, Barrett, A.J. (1987) *Trends Biol. Sci.* 12:13-196; Turk, V. and Bode, W. (1991) *FEBS Lett.* 285:213-219; Abrahamson, M. (1994) *Methods Enzymol.* 244:685-700). Cystatins bind to their target enzymes reversibly, in a one-to-one stoichiometry, with high affinity ($K_i=10^{-9}$

to 10^{-12} M). They display regulatory roles against the proteolytic activities of cysteine proteinases in physiological as well as pathological conditions.

All cystatins are members of one evolutionary superfamily (the cystatin superfamily) consisting of at least three distinct subfamilies of closely related proteins, referred to as stefins, cystatins and kininogens. Proteins of Family 1, the stefins, consist of about 100 amino acid residues with neither disulfide bonds nor carbohydrate groups. Proteins of Family 2, the cystatins, consist of about 115 amino acid residues which contain one or two disulfide loops near their C-terminus. Proteins of Family 3, the kininogens, are made up of three contiguous type-2 cystatin domains, followed by an additional domain of variable length containing a bradykinin sequence. The kininogens are multifunctional plasma glycoproteins involved in blood coagulation.

Members of the cystatin superfamily have been implicated in regulating tumor progression and metastatic potential (for a review see *e.g.*, Calkins, C.C. and Sloane, B.F. (1995) *Biol. Chem. Hoppe-Seyler*, 376:71-80). For example, cystatin C has been observed to be secreted from human colon carcinoma cell lines, as well as from a human fibrosarcoma cell line (Corticchiato, (1992) *Int. J. Cancer*, 52:645-652. Stefins markedly decreased the stimulated motility of both human melanoma cells and W256 carcinoma cells, implying that cysteine proteinases and their inhibitors may have a direct role in the development of a migratory response *per se* in tumor cells (Boike, G. *et al.* (1992) *Melanoma Res.* 1:333-340. Moreover, immunocytochemical localization of stefin A was greater in the fibrous stroma of less invasive breast cancer than in more invasive forms (Lah, T.T. *et al.* (1992) *Int. J. Cancer*, 50:36-44; Lah, T.T. *et al.* (1992) *Biol. Chem. Hoppe Seyler*, 373:595-604).

Members of the cystatin superfamily also have been implicated in the pathology of other disease conditions. For example, a mutant form of cystatin C is associated with the autosomal dominant disease, hereditary cystatin C amyloid angiopathy (HCCAA), and is deposited in amyloid fibrils of the patients, while native cystatin C was purified from cerebrospinal fluid (Ghiso, J. *et al.* (1986) *Proc. Natl. Acad. Sci. USA* 83:2974-2978; Olafsson, I. *et al.* (1990) *Scand. J. Clin. Lab. Invest.* 50:85-93).

Summary of the Invention

This invention provides an isolated nucleic acid molecule encoding a novel cysteine proteinase inhibitor, termed cystatin M, which is a member of the Family 2 cystatins. A partial cystatin M cDNA was originally identified by its differential expression in a primary mammary epithelial tumor cell line, as compared to a metastatic mammary tumor cell line derived from the same patient, using the differential display method. Subsequently, a full-length cDNA was isolated and identified as being down-regulated in the metastatic cells as compared to their parental cells from the primary tumor. These results indicate that the cystatin M gene is regulated at the transcriptional level, and its loss of expression is

associated with the metastatic phenotype, consistent with a tumor/metastasis suppressor function of the normal protein product along the metastatic cascade. The cystatin M protein displays approximately 40% sequence homology to human cystatins of Family 2 and exhibits cysteine proteinase inhibitory activity against the prototypical cysteine proteinase papain. In addition to its downregulation in metastatic mammary tumor cell lines, cystatin M mRNA is detectable in a variety of normal human tissues, undetectable in a variety of non-mammary tumor cell lines and markedly upregulated in replicatively senescent cells as compared to dividing or quiescent cells.

One aspect of the invention pertains to isolated nucleic acid molecules (*e.g.*, cDNAs) comprising a nucleotide sequence encoding cystatin M or biologically active portions thereof, as well as nucleic acid fragments suitable as hybridization probes for the detection of cystatin M-encoding nucleic acid (e.g., mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 1, or the coding region thereof, or encodes the amino acid sequence of SEQ ID NO: 2. In another embodiment, the isolated nucleic acid molecule encodes a protein which comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2 and inhibits the activity of papain *in vitro*. Preferably, the protein is at least 70 %, more preferably at least 80 %, even more preferably at least 90 % and most preferably at least 95 % homologous to the amino acid sequence of SEQ ID NO: 2. In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1. Preferably, the isolated nucleic acid corresponds to a naturally-occurring nucleic acid. More preferably, the isolated nucleic acid encodes naturally-occurring human cystatin M.

Moreover, given the disclosure herein of a cystatin M-encoding cDNA sequence (*e.g.*, SEQ ID NO: 1), antisense nucleic acid molecules (i.e, molecules which are complimentary to the coding strand of the cystatin M cDNA sequence) are also provided by the invention.

Another aspect of the invention pertains to recombinant expression vectors containing the nucleic acid molecules of the invention and host cells into which such recombinant expression vectors have been introduced. In one embodiment, such a host cell is used to produce cystatin M protein by culturing the host cell in a suitable medium. If desired, cystatin M protein can be then isolated from the medium or the host cell.

Yet another aspect of the invention pertains to transgenic non-human animals in which a cystatin M gene has been introduced or altered. In one embodiment, the genome of the nonhuman animal has been altered by introduction of a nucleic acid molecule of the invention encoding cystatin M as a transgene. In another embodiment, an endogenous cystatin M gene within the genome of the nonhuman animal has been altered, *e.g.*, functionally disrupted, by homologous recombination.

Still another aspect of the invention pertains to isolated cystatin M protein. The invention provides an isolated preparation of cystatin M. In preferred embodiments, the cystatin M protein comprises amino acids 1-149 of SEQ ID NO: 2 or about amino acids 22-149 of SEQ ID NO: 2 (lacking an amino-terminal signal sequence). In other embodiments, the isolated cystatin M protein comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2 and inhibits the activity of papain *in vitro*. Preferably, the protein is at least 70 %, more preferably at least 80 %, even more preferably at least 90 % and most preferably at least 95 % homologous to the amino acid sequence of SEQ ID NO: 2. In yet another embodiment, the cystatin M protein is glycosylated.

A cystatin M protein of the invention can be incorporated into a pharmaceutical composition comprising the protein and a pharmaceutically acceptable carrier. Moreover, the invention provides a fusion protein comprising a cystatin M polypeptide operatively linked to a non-cystatin M polypeptide.

The cystatin M proteins of the invention, or fragments thereof, can be used to prepare anti-cystatin M antibodies. The invention provides an antigenic peptide of cystatin M comprising at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO: 2 and encompassing an epitope of cystatin M such that an antibody raised against the peptide forms a specific immune complex with cystatin M. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues. The invention further provides an antibody that specifically binds cystatin M. In one embodiment, the antibody is monoclonal. In another embodiment, the antibody is coupled to a detectable substance. In yet another embodiment, the antibody is incorporated into a pharmaceutical composition comprising the antibody and a pharmaceutically acceptable carrier.

Another aspect of the invention pertains to methods for detecting the presence of cystatin M in a biological sample. In a preferred embodiment, the method involves contacting a biological sample (*e.g.*, a tumor sample) with an agent capable of detecting cystatin M protein or mRNA such that the presence of cystatin M is detected in the biological sample. The agent can be, for example, a labeled or labelable nucleic acid probe capable of hybridizing to cystatin M mRNA or a labeled or labelable antibody capable of binding to cystatin M protein. The invention further provides methods for diagnosis of a subject with a tumor based on detection of cystatin M protein or mRNA. In one embodiment, the method involves contacting a tumor sample from the subject with an agent capable of detecting cystatin M protein or mRNA, determining the amount of cystatin M protein or mRNA expressed in the tumor sample, comparing the amount of cystatin M protein or mRNA expressed in the tumor sample to a control sample and forming a diagnosis based on the amount of cystatin M protein or mRNA expressed in the tumor

sample as compared to the control sample. Preferably, the tumor sample is a mammary tumor sample. Kits for detecting cystatin M in a biological sample are also within the scope of the invention.

The cystatin M protein of the invention, and other agents related thereto, can be used to modulate the cystatin M cysteine proteinase inhibitory activity associated with a cell (*e.g.*, in the cell, secreted by the cell or in the extracellular milieu surrounding the cell). Accordingly, in one embodiment, the invention provides a method for stimulating the cystatin M cysteine proteinase inhibitory (CPI) activity associated with a cell by contacting the cell with an agent that stimulates cystatin M CPI activity. Such an agent can be, for example, an active cystatin M protein which is cultured with the cell or a nucleic acid encoding cystatin M that has been introduced into the cell. In a preferred embodiment, cystatin M CPI activity is stimulated in tumor cells, such as mammary tumor cells, in which endogenous cystatin M expression is low or absent. Alternatively, in another embodiment, the invention provides a method for inhibiting the cystatin M CPI activity associated with a cell by contacting the cell with an agent that inhibits cystatin M CPI activity. Such an agent can be, for example, an antisense cystatin M nucleic acid molecule or an anti-cystatin M antibody. The methods of the invention for modulating cystatin M activity can be applied *in vitro* (*e.g.*, with cells in culture) or *in vivo*, wherein an agent that modulates cystatin M CPI activity is administered to the subject. In a preferred embodiment, the invention provides a method for inhibiting development or progression of a metastatic phenotype in a tumor cell comprising contacting the tumor cell with an agent which elevates the amount of cystatin M in or around the tumor cell.

Screening methods for identifying modulators of the cystatin M expression or cystatin M cysteine proteinase inhibitory activity are also encompassed by the invention. In one embodiment, the modulator stimulates cystatin M expression or activity. In another embodiment, the modulator inhibits cystatin M expression or activity.

Brief Description of the Drawings

Figure 1 is the complete cDNA sequence and deduced amino acid sequence of human cystatin M (SEQ ID NOs: 1 and 2, respectively).

Figure 2 is a comparison of the amino acid sequence of cystatin M to other proteins of the human cystatin Family 2 (cystatins C, D, S, SN and SA) and chicken cystatin.

Figure 3 is a photograph of a Northern hybridization depicting the expression of cystatin M mRNA in normal mammary epithelial cell lines (76N, 70N), primary mammary epithelial tumor cell lines (21PT, 21NT) and metastatic mammary epithelial tumor cell lines (21MT-2, 3BT479, MCF7, T-47D, ZR-75-1, BT474, MDA361, MDA157, MDA231, MDA-MB-435, MDA-MB-436), as well as normal breast fibroblasts (56NF) and normal leukocytes (Leuk.).

Figure 4 is a photograph of a Northern hybridization depicting the expression of cystatin M mRNA in the following normal human tissues: heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas.

Figure 5A is a photograph of a Northern hybridization depicting the expression of cystatin M mRNA in passage 11 quiescents (days 2-12 post-plating) and in senescing cells (days 15-22 post-plating).

Figure 5B is a bar graph depicting the relative expression of cystatin M mRNA in quiescent and senescent cells.

Figure 6 is a photograph of an SDS-PAGE gel depicting lysate of bacteria cells transformed with pGEX-2T (lane 1) and glutathione-affinity-purified material therefrom (lane 2), lysate from bacterial cells transformed with pGEX-2T/cystatin M (lane 3) and glutathione-affinity-purified material therefrom (lane 4), and the GST-cystatin M fusion protein treated with thrombin for 0 minutes (lane 5), 2 minutes (lane 6), 30 minutes (lane 7) or 90 minutes (lane 8).

Figure 7A is a photograph of a Western blot depicting the expression of cystatin M protein in lysates (L) or supernatants (S) of a normal human mammary epithelial cell line (70N), a primary mammary epithelial tumor cell line (21PT) or malignant mammary epithelial tumor cell lines (MDA435, MDA157, BT549).

Figure 7B is a photograph of immunoprecipitates of culture supernatants of a primary mammary epithelial tumor cell line (21PT) or a malignant mammary epithelial tumor cell line (MDA435) with either preimmune serum or immune serum raised against recombinant cystatin M.

Figure 8A is a graph depicting the cleavage of Z-Phe-Arg-MCA by papain in the absence of any inhibitors ("no I") or in the presence of recombinant cystatin M fusion protein ("[I]=3nM").

Figure 8B is a Dixon plot graph for the estimation of the Ki value of recombinant cystatin M fusion protein for the inhibition of papain activity ("S" represents Z-Phe-Arg-MCA substrate: "FP" represents cystatin M fusion protein).

Figure 9 is a photograph of a Western blot depicting cystatin M protein immunoprecipitated from 21PT cell culture supernatant that was incubated in the absence (lane 1) or presence (lane 2) of N-Glycosidase F. Cystatin M cleaved from rGST-cystatin M fusion protein by thrombin is shown in lane 3. The arrow on the right indicates glycosylated cystatin M.

Detailed Description of the Invention

This invention pertains to a novel cysteine proteinase inhibitor (CPI) which is a member of the Family 2 cystatins. The CPI of the invention is named cystatin M but may also be referred to herein as 6A2, its clone designation. A partial cDNA encoding human cystatin M was originally isolated using the differential display method of Liang and

Pardee (*Science* (1992) 257:967-970) in experiments designed to identify mRNA transcripts differentially expressed in a primary mammary epithelial tumor cell line (21PT) as compared to a metastatic mammary tumor cell line (21MT-1) derived from the same patient. A full-length cDNA was subsequently isolated using the partial cDNA as a hybridization probe to screen a cDNA library prepared from the human mammary epithelial tumor cell line 21PT (described in further detail in Example 1). The nucleotide sequence of the isolated human cystatin M cDNA, and the predicted amino acid sequence of the human cystatin M protein, are shown in Figure 1 and in SEQ ID NOs: 1 and 2, respectively. As will be described further herein, cystatin M shares certain structural features with other members of the Family 2 human cystatins and chicken cystatin, but is only approximately 40% (or less) homologous to these proteins. Moreover, cystatin M mRNA has a distinct expression pattern that distinguishes it from other cystatins (described in further detail in Example 2). Cystatin M has been expressed as a recombinant protein (see Example 3) and used an immunogen to raise anti-cystatin M antibodies (see Example 4). Immunoprecipitation experiments with these antibodies demonstrated that 21PT cells express and secrete a native protein corresponding in size and immunoreactivity to the recombinant cystatin M (see Example 4). Moreover, recombinant cystatin M is effective at inhibiting the activity of the prototypical cysteine proteinase papain (see Example 5), demonstrating that cystatin M can function as a CPI. Approximately 30-40% of the native cystatin M protein occurs in a glycosylated form in a tumor cell line (see Example 6). Furthermore, transfection of a recombinant expression vector encoding cystatin M into a metastatic breast tumor cell line that does not express endogenous cystatin M leads to secretion of glycosylated recombinant cystatin M from the tumor cell line (see Example 7).

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode cystatin M or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify cystatin M-encoding nucleic acid (*e.g.*, cystatin M mRNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA). The nucleic acid molecule may be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated cystatin M nucleic acid molecule may contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the

nucleic acid is derived (*e.g.*, a human mammary epithelial cell). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, may be free of other cellular material.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO: 1. The sequence of SEQ ID NO: 1 corresponds to the human cystatin M cDNA. This cDNA comprises sequences encoding the cystatin M protein (*i.e.*, "the coding region", from nucleotides 24 to 470), as well as 5' untranslated sequences (nucleotides 1 to 23) and 3' untranslated sequences (nucleotides 471 to 598). Alternatively, the nucleic acid molecule may comprise only the coding region of SEQ ID NO: 1 (*e.g.*, nucleotides 24 to 470).

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of SEQ ID NO: 1, for example a fragment encoding a biologically active portion of cystatin M. The term "biologically active portion of cystatin M" is intended to include portions of cystatin M that retain the ability to inhibit cysteine proteinase activity. The ability of portions of cystatin M to inhibit cysteine proteinase activity can be determined in standard *in vitro* cysteine proteinase assays, for example using papain as the cysteine proteinase (described further below and in Example 5). In one embodiment, the biologically active portion of cystatin M is a mature form of cystatin M in which a hydrophobic, amino-terminal signal sequence (encompassing approximately amino acids 1-21) is absent. Accordingly, the mature form of cystatin M preferably comprises about amino acid residues 22 to 149 (*i.e.*, the nucleic acid molecule comprises nucleotides 87-470 of SEQ ID NO: 1). Leu-22 preferably is the N-terminal residue of the mature protein, although more than one native isoform differing in the length of the N-terminal sequence may exist for cystatin M, as has been reported for other cystatins. Consequently, the skilled artisan will appreciate that some flexibility exists in the N-terminus of the mature form of cystatin M lacking a signal sequence. For example, the mature form may begin with Ala-21 or Pro-23. Additional nucleic acid fragments encoding biologically active portions of cystatin M can be prepared by isolating a portion of SEQ ID NO: 1, expressing the encoded portion of cystatin M protein or peptide (*e.g.*, by recombinant expression *in vitro*) and assessing the cysteine proteinase inhibitory activity of the encoded portion of cystatin M protein or peptide.

The invention further encompasses nucleic acid molecules that differ from SEQ ID NO:1 (and portions thereof) due to degeneracy of the genetic code and thus encode the same cystatin M protein as that encoded by SEQ ID NO: 1. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO: 2. Moreover, the invention encompasses nucleic acid molecules that encode biologically active portions of SEQ ID NO: 2. For example, in one embodiment, the nucleic acid molecule encodes a portion of the amino acid sequence shown in SEQ ID NO: 2 corresponding to a mature form of cystatin M in which a hydrophobic, N-terminal signal sequence (*e.g.*, about amino

acid residues 1-21) is absent. In a preferred embodiment, this mature form encompasses about amino acid residues 22-149 of SEQ ID NO: 2.

A nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a human cystatin cDNA can be isolated from a normal mammary epithelial cell line cDNA library using all or portion of SEQ ID NO: 1 as a hybridization probe and standard hybridization techniques (*e.g.*, as described in Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO: 1 can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon the sequence of SEQ ID NO: 1. For example, mRNA can be isolated from normal mammary epithelial cells (*e.g.*, by the guanidinium-thiocyanate extraction procedure of Chirgwin *et al.* (1979) *Biochemistry* 18: 5294-5299) and cDNA can be prepared using reverse transcriptase (*e.g.*, Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for PCR amplification can be designed based upon the nucleotide sequence shown in SEQ ID NO: 1. For example, primers suitable for amplification of the segment of SEQ ID NO: 1 encoding amino acid residues 22 to 149 are shown in SEQ ID NOs: 3 and 4. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to cystatin M nucleotide sequence can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In addition to the human cystatin M nucleotide sequence shown in SEQ ID NO: 1, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of cystatin M may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the cystatin M gene may exist among individuals within a population due to natural allelic variation. Natural allelic variation has been observed with other members of the Family 2 cystatins. For example, two allelic variants of the cystatin D gene, resulting in an amino acid polymorphism at the protein level, have been described (Balbin, M. *et al.* (1993) *Hum. Genet.* 90:668-669; Balbin, M. *et al.* (1994) *J. Biol. Chem.* 269:23156-23162). Such natural allelic variations can typically result in 1-5 % variance in the nucleotide sequence of the a gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in cystatin M that are the result of natural allelic variation and that do not alter the functional activity of cystatin M are intended to be within the scope of the invention. Moreover, nucleic acid molecules

encoding cystatin M proteins from other species, and thus which have a nucleotide sequence which differs from the human sequence of SEQ ID NO: 1, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and nonhuman homologues of the human cystatin M cDNA of the invention can be isolated based on their homology to the human cystatin M nucleic acid disclosed herein using the human cDNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1. In other embodiment, the nucleic acid is at least 30, 50, 100, 250 or 500 nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60 % homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that at least sequences at least 65 %, more preferably at least 70 %, and even more preferably at least 75 % homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO: 1 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein). In on embodiment, the nucleic acid encodes a natural human cystatin M. In another embodiment, the nucleic acid molecule encodes a murine homologue of human cystatin M.

In addition to naturally-occurring allelic variants of the cystatin M sequence that may exist in the population, the skilled artisan will further appreciate that changes may be introduced by mutation into the nucleotide sequence of SEQ ID NO: 1, thereby leading to changes in the amino acid sequence of the encoded cystatin M protein, without altering the functional ability of the cystatin M protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues may be made in the sequence of SEQ ID NO: 1. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of cystatin (*e.g.*, the sequence of SEQ ID NO: 2) without altering the cysteine proteinase inhibitory activity of cystatin, whereas an "essential" amino acid residue is required for CPI activity. Amino acid residues of cystatin M that are strongly conserved among members of the Family 2 cystatins (*e.g.*, conserved in 6 of 7 of the family members whose amino acid sequences are aligned and compared in

Figure 2) are predicted to be essential in cystatin M and thus are not likely to be amenable to alteration. For example, all members of the Family 2 cystatins, including cystatin M, characteristically contain four cysteine residues that participate in the formation of two intrachain disulfide bridges. Reduction of both disulfide bonds of chicken cystatin destroys the cysteine proteinase binding ability of the protein (Björk, I. and Ylinenjärvi, K. (1992) *Biochemistry* 31:8597-8602). Accordingly, these conserved cysteine residues in cystatin M (cys-98, cys-113, cys-126 and cys-146) are not likely to be amenable to mutation.

Moreover, structure/function and crystallographic analyses of various members of the Family 2 cystatins have identified other residues and/or regions that are important for the CPI activity of these cystatins (see *e.g.*, Abrahamson, M. *et al.* (1987) *J. Biol. Chem.* 262:9688-9694; Bode, W. *et al.* (1988) *EMBO J.* 7:2593-2599; Lindahl, P. *et al.* (1992) *Biochem. J.* 286:165-171; Hall, A. *et al.* (1993) *Biochem. J.* 291:123-129; Engh, R.A. *et al.* (1993) *J. Mol. Biol.* 234:1060-1069; Dieckmann, T. *et al.* (1993) *J. Mol. Biol.* 234:1048-1059; Bobek, L.A. *et al.* (1994) *Gene* 151:303-308; Hall, A. *et al.* (1995) *J. Biol. Chem.* 270:5115-5121; and Björk, I. *et al.* (1995) *Biochem. J.* 306:513-518). Using chicken cystatin and papain as a model inhibitor-enzyme system, it appears that the enzyme-inhibitor complex is formed mainly by hydrophobic interactions between cystatin and papain, with contributions from the N-terminal segment of chicken cystatin (Leu8-Gly9 of the mature protein) in a substrate-like interaction with the active-site cleft of papain, as well as the first hairpin loop (Glu53-Gly57), a conserved region in all these inhibitors, and the second hairpin loop (Pro103-Leu105). The N-terminal segments of cystatin C and chicken cystatin interact in a substrate-like manner with the subsites S3 and S2 of the proteinase. For example, mutagenesis studies have demonstrated the importance of Gly11 in human cystatin C (corresponding to Gly9 in chicken cystatin) in the ability of the protein to adopt a conformation suitable for interaction with the substrate-binding pockets of cysteine proteinases (Hall, A. *et al.* (1993) *Biochem. J.* 291:123-129). However, mutagenesis of residues outside the N-terminal region and the first and second hairpin loops may be inconsequential for the CPI activity of the protein. For example, an Arg to Trp substitution made at position 117 of human cystatin S had no effect on the inhibitory activity of the protein (Saitoh, E. and Isemurà, S. (1994) J. Biochem. 116:399-405).

Cystatin M, like all other members of the family, shares all three conserved domains important for CPI activity, namely the N-terminal region and the first and second hairpin loops. In particular, Gly-36 of SEQ ID NO: 2, near the N-terminal domain of the cystatin M preprotein, corresponds to the conserved Gly-11 of mature cystatin C and Gly-9 of mature chicken cystatin. Cystatin M also contains the first and second hairpin loop motifs associated with cysteine proteinase inhibitory activity: the 'QXVXG' motif in the middle of the molecule (QLVAG at positions 80 to 84 of cystatin M shown in SEQ ID NO: 2; QIVAG in cystatin C, QLVSG in chicken cystatin), as well as the VPW sequence near the carboxy terminal (positions 133 to 135 of SEQ ID NO: 2), which is also conserved in all

known cystatins. Thus, these highly conserved regions in cystatin M necessary for the CPI activity of the protein are not likely to be amenable to mutation. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved among members of the Family 2 cystatins) may not be essential for CPI activity and thus are likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding cystatin M proteins that contain changes in amino acid residues that are not essential for CPI activity , *e.g.*, residues that are not conserved or only semi-conserved among members of the Family 2 cystatins. Such cystatin M proteins differ in amino acid sequence from SEQ ID NO: 2 yet retain CPI activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2 inhibits the activity of papain *in vitro*. Preferably, the protein encoded by the nucleic acid molecule is at least 70 % homologous to SEQ ID NO: 2, more preferably at least 80 % homologous to SEQ ID NO: 2, even more preferably at least 90 % homologous to SEQ ID NO: 2, and most preferably at least 95 % homologous to SEQ ID NO: 2.

To determine the percent homology of two amino acid sequences (*e.g.*, SEQ ID NO: 2 and a mutant form thereof), the sequences are aligned for optimal comparison purposes (*e.g.*, gaps may be introduced in the sequence of one protein for optimal alignment with the other protein). The amino acid residues at corresponding amino acid positions are then compared. When a position in one sequence (*e.g.*, SEQ ID NO: 2) is occupied by the same amino acid residue as the corresponding position in the other sequence (*e.g.*, a mutant form of cystatin M), then the molecules are homologous at that position (*i.e.*, as used herein amino acid "homology" is equivalent to amino acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*,

% homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding a cystatin M protein homologous to the protein of SEQ ID NO: 2 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 1 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO: 1 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art, including basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid),

uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in cystatin M is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a cystatin M coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for proteolytic activity to identify mutants that retain proteolytic activity. Following mutagenesis of SEQ ID NO: 1, the encoded protein can be expressed recombinantly (*e.g.*, as described in Example 3) and the cysteine proteinase inhibitory activity of the protein can be determined.

A suitable assays for testing the cysteine proteinase inhibitory activity of portions of cystatin M proteins and mutated cystatin M proteins is described in detail in Example 5. Briefly, a recombinant cystatin M protein (*e.g.*, a mutated or truncated form of SEQ ID NO: 2) is incubated with papain (commercially available from Boehringer Mannheim) and a synthetic substrate for papain, such as the fluorogenic substrate Z-Phe-Arg-MCA (commercially available from Sigma Chemical Co., St. Louis, MO). The amount of 7-amino-4-methylcoumarin liberated from the synthetic substrate is then determined fluorometrically as a measure of the cysteine proteinase activity of papain. To determine the inhibitory effect of the cystatin M protein, the papain activity in the presence of the cystatin M protein is compared with the papain activity in the absence of the cystatin M protein.

In addition to the nucleic acid molecules encoding cystatin M proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid.

The antisense nucleic acid can be complementary to an entire cystatin M coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding cystatin M. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the entire coding region of SEQ ID NO: 1 comprises nucleotides 24-470). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding cystatin M. The term "noncoding region" refers to 5' and 3'

sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding cystatin M disclosed herein (*e.g.*, SEQ ID NO: 1), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule may be complementary to the entire coding region of cystatin M mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of cystatin M mRNA. For example, the antisense oligonucleotide may be complementary to the region surrounding the translation start site of cystatin M mRNA. An antisense oligonucleotide can be, for example, about 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

In another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. A ribozyme having specificity for a cystatin M-encoding nucleic acid can be designed based upon the nucleotide sequence of a cystatin M cDNA disclosed herein (*i.e.*, SEQ ID NO: 1). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the base sequence of the active site is complementary to the base sequence to be cleaved in a cystatin M-encoding mRNA. See for example Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, cystatin M mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See for example Bartel, D. and Szostak, J.W. (1993) *Science* 261: 1411-1418.

II. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding cystatin M (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid",

which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, cystatin M proteins, mutant forms of cystatin M, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of cystatin M in prokaryotic or eukaryotic cells. For example, cystatin M can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast

cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector may be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promotors directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc.; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In a preferred embodiment, exemplified in Example 3, the coding sequence of the mature form of cystatin M (*i.e.*, encompassing amino acids 22-149) is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-cystatin M. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant cystatin M unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that

the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nuc. Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the cystatin M expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari. *et al.*, (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA).

Alternatively, cystatin M can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.*, (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B., (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987), *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense

to cystatin mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to recombinant host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact; be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell may be any prokaryotic or eukaryotic cell. For example, cystatin M protein may be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker may be introduced into a host cell on the same vector as that encoding cystatin M or may be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection

(*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) cystatin M protein. Accordingly, the invention further provides methods for producing cystatin M protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding cystatin M has been introduced) in a suitable medium until cystatin M is produced. In another embodiment, the method further comprises isolating cystatin M from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which cystatin M-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous cystatin M sequences have been introduced into their genome or homologous recombinant animals in which endogenous cystatin M sequences have been altered. Such animals are useful for studying the function and/or activity of cystatin M and for identifying and/or evaluating modulators of cystatin M activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which one or more of the cells of the animal includes a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous cystatin M gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing cystatin M-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The human cystatin M cDNA sequence of SEQ ID NO: 1 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human cystatin M gene, such as a mouse cystatin M gene, can be isolated based on hybridization to the human cystatin cDNA (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the cystatin M transgene to direct expression of cystatin M protein to particular cells. Methods for generating transgenic animals via

embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the cystatin M transgene in its genome and/or expression of cystatin M mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding cystatin M can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a cystatin M gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the cystatin M gene. The cystatin M gene may be a human gene (*e.g.*, from a human genomic clone isolated from a human genomic library screened with the cDNA of SEQ ID NO: 1), but more preferably, is a non-human homologue of a human cystatin M gene. For example, a mouse cystatin M gene can be isolated from a mouse genomic DNA library using the human cystatin M cDNA of SEQ ID NO: 1 as a probe. The mouse cystatin M gene then can be used to construct a homologous recombination vector suitable for altering an endogenous cystatin M gene in the mouse genome. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous cystatin M gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous cystatin M gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous cystatin M protein). In the homologous recombination vector, the altered portion of the cystatin M gene is flanked at its 5' and 3' ends by additional nucleic acid of the cystatin M gene to allow for homologous recombination to occur between the exogenous cystatin M gene carried by the vector and an endogenous cystatin gene in an embryonic stem cell. The additional flanking cystatin M nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see *e.g.*, Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced cystatin gene has homologously recombined with the endogenous cystatin M gene are selected (see *e.g.*, Li, E. *et al.* (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see *e.g.*, Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp.

113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

III. Isolated Cystatin M Proteins and Anti-Cystatin M Antibodies

Another aspect of the invention pertains to isolated cystatin M proteins, and biologically active portions thereof, as well as peptide fragments suitable as immunogens to raise anti-cystatin M antibodies. The invention provides an isolated preparation of cystatin M, or a biologically active portion thereof. An "isolated" protein is substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. In a preferred embodiment, the cystatin M protein has an amino acid sequence shown in SEQ ID NO: 2. In other embodiments, the cystatin M protein is substantially homologous to SEQ ID NO: 2 and retains the functional activity of the protein of SEQ ID NO: 2 yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the cystatin M protein is a protein which comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2 and inhibits the activity of papain *in vitro*. Preferably, the protein is at least 70 % homologous to SEQ ID NO: 2, more preferably at least 80 % homologous to SEQ ID NO: 2, even more preferably at least 90 % homologous to SEQ ID NO: 2, and most preferably at least 95 % homologous to SEQ ID NO: 2.

An isolated cystatin M protein may comprise the entire amino acid sequence of SEQ ID NO: 2 (*i.e.*, amino acids 1-149) or a biologically active portion thereof. For example, a biologically active portion of cystatin M can comprise a mature form of cystatin M in which a hydrophobic, amino-terminal signal sequence is absent. In one embodiment, such a mature form of cystatin M comprises about amino acids 22-149 of SEQ ID NO: 2. The term "about amino acids 22-149" is intended to indicate that there is some flexibility in the amino-terminal residue, as discussed further in subsection I above. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for cysteine proteinase inhibitory activity as described in detail above.

In one embodiment, the cystatin M protein of the invention is glycosylated. Native cystatin M has been found to exist both as a 14.5 kDa form and as a 20-22 kDa form, the

latter representing a glycosylated form. Recombinant expression of cystatin M protein in mammalian cells (*e.g.*, a metastatic breast tumor cell line) can lead to production of the glycosylated form of cystatin M.

Cystatin M proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the cystatin M protein is expressed in the host cell. The cystatin M protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, a cystatin M protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native cystatin M protein can be isolated from cells (*e.g.*, cultured human mammary epithelial cells), for example using an anti-cystatin M antibody (discussed further below).

The invention also provides cystatin M fusion proteins. As used herein, a cystatin M "fusion protein" comprises a cystatin M polypeptide operatively linked to a non-cystatin M polypeptide. A "cystatin M polypeptide" refers to a polypeptide having an amino acid sequence corresponding to cystatin M, whereas a "non-cystatin M polypeptide" refers to a polypeptide having an amino acid sequence corresponding to another protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the cystatin M polypeptide and the non-cystatin M polypeptide are fused in-frame to each other. The non-cystatin M polypeptide may be fused to the N-terminus or C-terminus of the cystatin M polypeptide. For example, in one embodiment the fusion protein is a GST-cystatin M fusion protein in which the cystatin M sequences are fused to the C-terminus of the GST sequences (see Example 3). Such fusion proteins can facilitate the purification of recombinant cystatin M. In another embodiment, the fusion protein is a cystatin M protein containing a heterologous signal sequence at its N-terminus. For example, the native cystatin M signal sequence (i.e, about amino acids 1-21) can be removed and replaced with a signal sequence from another protein. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of cystatin M may be increased through use of a heterologous signal sequence.

Preferably, a cystatin M fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary

overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A cystatin M-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the cystatin M protein.

An isolated cystatin M protein, or fragment thereof, can be used as an immunogen to generate antibodies that bind cystatin M using standard techniques for polyclonal and monoclonal antibody preparation. The full-length cystatin M protein can be used or, alternatively, the invention provides antigenic peptide fragments of cystatin M for use as immunogens. The antigenic peptide of cystatin M comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO: 2 and encompasses an epitope of cystatin M such that an antibody raised against the peptide forms a specific immune complex with cystatin M. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of cystatin M that are located on the surface of the protein, *e.g.*, hydrophilic regions. A hydrophobicity analysis of the cystatin M protein sequence indicates three hydrophilic regions that are preferred for use as antigenic peptides: amino acid residues 22-49 (corresponding to the amino-terminus of the mature cystatin M protein), amino acid residues 90-104 and amino acid residues 112-126 of SEQ ID NO: 2.

A cystatin M immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for examples, recombinantly expressed cystatin M protein or a chemically synthesized cystatin M peptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic cystatin M preparation induces a polyclonal anti-cystatin M antibody response.

Accordingly, another aspect of the invention pertains to anti-cystatin M antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as cystatin M. The invention provides polyclonal and monoclonal antibodies that bind cystatin M. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of cystatin M. A monoclonal antibody composition thus typically displays a single binding affinity for a particular cystatin M protein with which it immunoreacts.

Polyclonal anti-cystatin M antibodies can be prepared as described above by immunizing a suitable subject with a cystatin M immunogen (see also Example 4). The anti-cystatin M antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized cystatin M. If desired, the antibody molecules directed against cystatin M can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-cystatin M antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975, *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol* 127:539-46; Brown *et al.* (1980) *J Biol Chem* 255:4980-83; Yeh *et al.* (1976) *PNAS* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter *et al.* (1977) *Somatic Cell Genet.*, 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a cystatin M immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds cystatin M.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-cystatin M monoclonal antibody (see, *e.g.*, G. Galfre *et al.* (1977) *Nature* 266:55052; Gefter *et al.* *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinary skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines may be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from the American Type Culture Collection (ATCC), Rockville, Md. Typically, HAT-sensitive

mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed).

Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind cystatin M, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-cystatin M antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with cystatin M to thereby isolate immunoglobulin library members that bind cystatin M. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP*™ *Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* International Publication No. WO 92/18619; Dower *et al.* International Publication No. WO 91/17271; Winter *et al.* International Publication WO 92/20791; Markland *et al.* International Publication No. WO 92/15679; Breitling *et al.* International Publication WO 93/01288; McCafferty *et al.* International Publication No. WO 92/01047; Garrard *et al.* International Publication No. WO 92/09690; Ladner *et al.* International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum Antibod Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J Mol Biol* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *PNAS* 89:3576-3580; Garrad *et al.* (1991) *Bio/Technology* 9:1373-1377; Hoogenboom *et al.* (1991) *Nuc Acid Res* 19:4133-4137; Barbas *et al.* (1991) *PNAS* 88:7978-7982; and McCafferty *et al. Nature* (1990) 348:552-554.

Additionally, recombinant anti-cystatin M antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Patent Publication PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT Application WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *PNAS* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *PNAS*

84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl Cancer Inst.* 80:1553-1559); Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

An anti-cystatin M antibody (*e.g.*, monoclonal antibody) can be used to isolate cystatin M by standard techniques, such as affinity chromatography or immunoprecipitation (see *e.g.*, Example 4). An anti-cystatin M antibody can facilitate the purification of natural cystatin M from cells and of recombinantly produced cystatin M expressed in host cells. Moreover, an anti-cystatin M antibody can be used to detect cystatin M protein (*e.g.*, in a cellular lysate or cell supernatant). Detection may be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

IV. Pharmaceutical Compositions

The cystatin M proteins and anti-cystatin M antibodies of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the protein or antibody and a pharmaceutically acceptable carrier. As used herein the term "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. For example, solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates

or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a cystatin protein or anti-cystatin antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The

tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These may be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

V. Uses and Methods of the Invention

As described in more detail in Example 5, the cystatin M protein of the invention exhibits cysteine proteinase inhibitory activity. Accordingly, cystatin M is useful as a cysteine proteinase inhibitor, either *in vitro* or *in vivo*. The isolated nucleic acid molecules of the invention can be used to express cystatin M protein (*e.g.*, via a recombinant expression vector in a host cell), to detect cystatin M mRNA (*e.g.*, in a biological sample) and to modulate cystatin M activity, as discussed further below. Moreover, the anti-cystatin M antibodies of the invention can be used to detect and isolate cystatin M protein and modulate cystatin M activity, also discussed further below.

The invention provides a method for detecting the presence of cystatin M in a biological sample. The method involves contacting the biological sample with an agent capable of detecting cystatin M protein or mRNA such that the presence of cystatin M is

detected in the biological sample. A preferred agent for detecting cystatin M mRNA is a labeled or labelable nucleic acid probe capable of hybridizing to cystatin M mRNA. The nucleic acid probe can be, for example, the full-length cystatin M cDNA of SEQ ID NO: 1, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to cystatin M mRNA. A preferred agent for detecting cystatin M protein is a labeled or labelable antibody capable of binding to cystatin M protein. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled or labelable", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect cystatin M mRNA or protein in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of cystatin M mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of cystatin M protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Alternatively, cystatin M protein can be detected *in vivo* in a subject by introducing into the subject a labeled anti-cystatin M antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In a preferred embodiment of the detection method, the biological sample is a tumor sample. The tumor sample may comprise tumor tissue or a suspension of tumor cells. A tissue section, for example, a freeze-dried or fresh frozen section of tumor tissue removed from a patient, can be used as the tumor sample. Moreover, the tumor sample may comprise a biological fluid obtained from a tumor-bearing subject. Since cystatin M contains a signal sequence and is detectable in culture supernatants of a primary mammary epithelial tumor cell line (see the Examples), the protein is thought to be secreted and thus is likely to be detectable in biological fluids. Following collection, tumor samples can be stored at temperatures below $-20^{\circ}C$ to prevent degradation until the detection method is to be performed. A preferred tumor sample in which cystatin M mRNA or protein is to be detected is a mammary tumor sample.

The detection methods of the invention described above can be used as the basis for a method of diagnosis of a subject with a tumor. As described in further detail in Example 2, the expression pattern of cystatin M mRNA can differ between normal cells and

malignant cells and between primary tumor cells and metastatic tumor cells. For example, cystatin M mRNA levels are detectable in several normal mammary epithelial cell lines, elevated in several primary mammary epithelial tumor cell lines and undetectable in several metastatic mammary epithelial tumor cell lines. Immunoprecipitation experiments (see Example 4) indicate that the cystatin M protein expression pattern mimics the mRNA expression pattern. Additional experiments (see Example 2) indicate that cystatin M mRNA is expressed in many normal non-mammary tissues (such as lung, pancreas, ovaries and prostate) and undetectable in many malignant non-mammary tissues (including tumor cells from lung, pancreas, ovaries and prostate). Accordingly, the invention provides a diagnostic method comprising:

contacting a tumor sample from a subject with an agent capable of detecting cystatin M protein or mRNA;

determining the amount of cystatin M protein or mRNA expressed in the tumor sample;

comparing the amount of cystatin M protein or mRNA expressed in the tumor sample to a control sample; and

forming a diagnosis based on the amount of cystatin M protein or mRNA expressed in the tumor sample as compared to the control sample.

In one embodiment, the control is from normal cells and the tumor sample is a suspected primary tumor sample. Primary malignancy of the tumor cell sample can be diagnosed based on an increase in the level of expression of cystatin M mRNA or protein in the tumor sample as compared to the control. In another embodiment, the control is from normal cells or a primary tumor and the tumor sample is a suspected metastatic tumor sample. Acquisition of the metastatic phenotype by the suspected metastatic tumor sample can be diagnosed based on a decrease in the level of, or absence of, cystatin M mRNA or protein in the tumor sample compared to the control.

The invention also encompasses kits for detecting the presence of cystatin M in a biological sample (*e.g.*, a tumor sample). For example, the kit can comprise a labeled or labelable agent capable of detecting cystatin M protein or mRNA in a biological sample; means for determining the amount of cystatin M in the sample; and means for comparing the amount of cystatin M in the sample with a standard. The agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect cystatin M mRNA or protein.

Another aspect of the invention pertains to methods of modulating cystatin M cysteine proteinase inhibitory activity associated with a cell, *e.g.*, for therapeutic purposes. Cystatin M cysteine proteinase inhibitory (CPI) activity "associated with a cell" is intended to include cystatin M CPI activity within the cell, secreted by the cell and in the extracellular milieu surrounding the cell. The modulatory method of the invention involves contacting the cell with an agent that modulates cystatin M cysteine proteinase inhibitory

(CPI) activity associated with the cell. In one embodiment, the agent stimulates cystatin M cysteine proteinase inhibitory activity. Examples of such stimulatory agents include active cystatin M protein and a nucleic acid molecule encoding cystatin M that has been introduced into the cell. In another embodiment, the agent inhibits the cystatin M CPI activity . Examples of such inhibitory agents include antisense cystatin M nucleic acid molecules and anti-cystatin M antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject).

Stimulation of cystatin M CPI activity is desirable in situations in which cystatin M is abnormally downregulated and/or in which increased cystatin M activity is likely to have a beneficial effect. One example of such a situation is in tumor cells, and in particular in inhibiting or preventing tumor cell metastasis. As demonstrated in Example 2, acquisition of a metastatic phenotype by tumor cells is associated with downregulation of cystatin M expression. Thus, increasing the expression and/or activity of cystatin M in or around the tumor cells is expected to inhibit the development or progression of the metastatic phenotype. Accordingly, in a specific embodiment, the invention provides a method for inhibiting development or progression of a metastatic phenotype in a tumor cell comprising contacting the tumor cell with an agent which elevates the amount of cystatin M in or around the tumor cell. The term "in or around the tumor cell" is intended to include cystatin M within the cell, secreted by the cell and in the extracellular milieu surrounding the cell. The agent that elevates cystatin M in or around the tumor cell can be cystatin M protein itself. For example, since cystatin M is a secreted protein, it is likely that it exerts tumor suppressive effects extracellularly. Thus, cystatin M, preferably in a pharmaceutically acceptable carrier, can be administered to a tumor-bearing subject by an appropriate route to inhibit the development or progression of the metastatic phenotype of the tumor. Suitable routes of administration include intravenous, intramuscular or subcutaneous injection, injection directly into the tumor site or implantation of a device containing a slow-release formulation. The cystatin M preparation can also be incorporated into liposomes or other carrier vehicles to facilitate delivery to the tumor site. A non-limiting dosage range is 0.001 to 100 mg/kg/day, with the most beneficial range to be determined by routine pharmacological methods.

Alternative to administration of cystatin M protein itself, the development or progression of the metastatic phenotype can be inhibited in tumor cells by modifying them to express cystatin M by introducing into the tumor cells a nucleic acid encoding cystatin M (*e.g.*, via a recombinant expression vector). Expression vectors suitable for gene therapy, including retroviral and adenoviral vectors carrying appropriate regulatory elements, can be used to deliver the cystatin M-encoding nucleic acid to the tumor cells.

The ability of cystatin M protein or DNA to inhibit tumor progression and/or metastasis can be evaluated using *in vivo* and *in vitro* assays known in the art. For

example, a suitable *in vivo* assay utilizes the mammary epithelial tumor cell line MDA-MB-435, which forms tumors at the site of orthotopic injection and metastasizes in nude mice (described further in Price *et al.* (1990) *Cancer Res.* 50:717). MDA-MB-435 cells, which do not express detectable cystatin M mRNA, can be transfected with a cystatin M expression vector and stable transfectants can be selected (described further in Example 7). These transfectants can then be injected into nude mice. At 10-weeks post-inoculation, the mice are sacrificed and their tumors are excised and weighed to determine the effect of cystatin M expression on tumor progression and metastasis. A suitable *in vitro* assay is tumor cell invasion through reconstituted basement membrane matrix (*e.g.*, Matrigel) as described in Hendrix *et al.* (1987) *Cancer Letters* 38:137. The invasive ability of cystatin M-transfected MDA-MB-435 cells can be compared to untransfected MDA-MB-435 cells to determine the effect of cystatin M expression on tumor invasiveness.

In addition to tumor therapy, there are other situations in which stimulating cystatin M CPI activity may be desirable. Other members of the Family 2 cystatins, or portions thereof, have been shown to have anti-bacterial and/or anti-viral activity. For example, cystatin C and a tripeptide derivative thereof have been shown to inhibit herpes simplex virus replication (Björck, L *et al.* (1990) *J. Virol.* 64:941-943). Moreover, the same tripeptide derivative has been shown to block the growth of several bacterial strains (Björck, L *et al.* (1989) *Nature* 337:385-386). Antiviral effects have also been reported for chicken cystatin C (see *e.g.*, U.S. Patent No. 4,902,509 and EP 188 262, both by Turk *et al.* and U.S. Patent No. 5,124,443 by Bird *et al.*). Accordingly, cystatin M may be useful as an anti-bacterial and/or anti-viral agent. Cystatins have also been reported to have a therapeutically beneficial effect in periodontal disease (see *e.g.*, Lah, T.T. (1993) *J. Periodontol.* 64:485-491), in reducing dental caries (see *e.g.*, PCT International Publication No. WO 94/15578 by Revis *et al.*) and in the treatment of gastrointestinal ulcers (see *e.g.*, U.S. Patent No. 4,891,356 and PCT International Publication No. WO 89/00426, both by Szabo). Accordingly, cystatin M, administered as a mouthwash or oral formulation, may be useful in treating gingivitis, dental caries or gastrointestinal ulcers. Cystatins also have been reported to protect plants from various pests (*e.g.*, parasites, helminths, protozoans) (see *e.g.*, PCT International Publication No. WO 95/23229 by Atkinson *et al.* and EP 348 348 by Fowler *et al.*). Accordingly, cystatin M also may have agricultural use in combating plant pests.

In contrast to the foregoing situations in which stimulation of cystatin M CPI activity is desirable, there are other situations in which it may be desirable to decrease cystatin M activity using an inhibitory method of the invention. For example, as demonstrated in Example 2, cystatin mRNA expression is markedly upregulated in senescent cells. Thus, inhibiting the expression or activity of cystatin M in cells may be useful for inhibiting or delaying the onset of senescence in the cells. For example, the *in vitro* growth of particular cells which normally undergo senescence in culture could be

maintained and prolonged by the use of cystatin M inhibitory agents to inhibit or delay the onset of senescence in the cells.

The invention still further provides a method for identifying modulators of cystatin M expression or cysteine proteinase inhibitory (CPI) activity. In one embodiment, modulators of cystatin M CPI activity are identified in a method wherein cystatin M, a cysteine proteinase, a substrate for the cysteine proteinase and a test substance are incubated under conditions suitable for the cysteine proteinase to cleave the substrate. Cleavage of the substrate is then measured and the amount of cleavage of the substrate in the presence of the test substance is compared to the amount of cleavage of the substrate in the absence of the test substance. The test substance can then be identified as a modulator of cystatin M CPI activity based on this comparison. For example, when the amount of cleavage of the substrate in the presence of the test substance is less than the amount of cleavage of the substrate in the absence of the test substance, the test substance can thereby be identified as a stimulator of the CPI activity of cystatin M. Alternatively, when the amount of cleavage of the substrate in the presence of the test substance is greater than the amount of cleavage of the substrate in the absence of the test substance, the test substance can thereby be identified as an inhibitor of the cysteine proteinase inhibitory activity of cystatin M. A preferred cysteine proteinase for use in the method is papain. A preferred substrate for papain is the fluorogenic synthetic substrate Z-Phe-Arg-MCA.

In another embodiment, modulators of cystatin M expression are identified in a method wherein a cell is contacted with a test substance and the expression of cystatin M mRNA or protein in the cell is determined. The level of expression of cystatin M mRNA or protein in the presence of the test substance is compared to the level of expression of cystatin M mRNA or protein in the absence of the test substance. The test substance can then be identified as a modulator of cystatin M expression based on this comparison. For example, when expression of cystatin mRNA or protein is greater in the presence of the test substance than in its absence, the test substance is identified as a stimulator of cystatin M mRNA or protein expression. Alternatively, when expression of cystatin mRNA or protein is less in the presence of the test substance than in its absence, the test substance is identified as an inhibitor of cystatin M mRNA or protein expression. The level of cystatin M mRNA or protein expression in the cells can be determined by methods described above for detecting cystatin M mRNA or protein.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

EXAMPLE 1: Isolation and Characterization of Cystatin M cDNA

In this example, a partial cDNA encoding human cystatin M was first isolated by differential expression cloning using the differential display (DD) method. The partial cDNA was used as a hybridization probe in Northern blots to confirm the differential expression of cystatin M mRNA in a primary mammary epithelial tumor cell line as compared to a metastatic cell line derived from the same patient. The partial cDNA was then used as a probe to isolate a full-length cDNA, the sequence of which was then determined and analyzed.

Messenger RNAs expressed by a primary mammary epithelial tumor cell line, 21PT, and a metastatic cell line, 21MT-1, derived from the same patient were compared by DD. These cell lines are described further in Band, V. *et al.* (1990) *Cancer Res.* 50:7351-7357. The differential display method is described further in Liang, L. and Pardee, A.B. (1992) *Science* 257:967-970 and in Sager, R. *et al.* (1993) *FASEB J.* 7:964-970. Total cellular RNA was isolated from exponentially growing cultures of the 21PT and 21MT-1 cell lines by standard techniques. 50 µg of total cellular RNA from each sample was treated with DNase I in the presence of RNasin ribonuclease inhibitor, in order to remove any residual DNA contamination. Then, RNA was extracted with phenol/chloroform, precipitated with ethanol and redissolved in DEPC-treated water. Purified total RNA was subsequently reverse transcribed using a 3'-anchoring primer $T_{12}MA$ (where M is degenerate for G, C, or A). The resultant partial cDNAs were amplified by PCR using $T_{12}MA$ as the 3' primer and an arbitrary 10-mer, OM6 (Operon Technologies, Inc.) as the 5' primer in the presence of [^{35}S]dATP. The PCR products were compared side-by-side on a 6% acrylamide sequencing gel as ^{35}S-labelled partial cDNA fragments corresponding to the 3'-end of the mRNAs. Each lane contained 50-100 bands, most of which are identical in size and intensity between the two cell populations. A small number of bands (~1-2%) appeared in only one of the cell lines. In particular, one differentially displayed cDNA of about 0.3 kb, named 6A2, was seen in the primary tumor cells 21PT but was absent in the metastatic tumor cells 21MT-1. The band corresponding to 6A2 was recovered from the dried gel, reamplified by PCR, ^{32}P-labeled by the oligo-labeling method (Freinberg, A.P., and Vogelstein, B.A. (1983) *Anal. Biochem.* 132: 6) and used as a probe for hybridization of Northern blots containing RNA from 21PT and 21MT-1. A transcript of 0.6 kb was differentially expressed in 21PT as compared to 21MT-1.

Since the 6A2 PCR product gave a confirmatory differential signal by Northern blot, the 6A2 partial cDNA obtained from DD was reamplified by PCR, cloned into the PCRII vector using the TA cloning system (Invitrogen) and sequenced on both strands with T7 and SP6 primers. A cDNA library from the human mammary epithelial tumor cell line 21PT, constructed in Lambda Zap II (Stratagene, San Diego, CA) was screened according to standard cDNA library screening methods using the cloned PCR product as a probe. Several full-length cDNA clones were isolated. Subsequently, DNA was isolated from

recombinant phage clones, the phage insert was prepared by restriction endonuclease digestion, and was hybridized to total RNAs from normal and tumor cell lines. All clones tested displayed confirmatory differential expression on Northern blots as on the DD gel. Three clones were sequenced on both strands using an ABI automated sequencer, Model 373A, and the obtained nucleotide sequences were compared for verification. All three clones contained an ATG initiation codon at the 5'-region. The longest clone, 6A2-17, was selected for further sequence analysis.

The full-length nucleotide sequence and deduced amino acid sequence of the cystatin M clone 6A2-17 are shown in Figure 1 and in SEQ ID NOs: 1 and 2, respectively. Assuming that translation starts at the first ATG codon which lies in a Kozak consensus sequence, the full-length 6A2-17 cDNA clone contains an open reading frame of consisting of 447 nucleotides (24-470), a short 5'-untranslated sequence (1-23), and a 3'-untranslated sequence of 128 nucleotides, with a polyadenylation signal (AATAAA) (552-557) and a poly A tail. The partial 6A2 cDNA originally obtained from DD corresponds to nt 299-598. The amino acid sequence inferred from the nucleotide sequence of the full-length cDNA encodes a protein of 149 amino acid residues long with four cysteine residues towards its carboxy terminal domain. The first ATG codon is probably the major transcription start site, since the translated sequence aligns optimally with other human cystatins. In contrast, internal ATGs do not lie in a fair Kozak consensus.

The BLAST algorithm via Autosearch was used for nucleic acid sequence comparisons. Protein sequence comparisons were performed on GCG with final alignments on PILEUP and PRETTYPLOT (Altschul, S.F. *et al.* (1990) *J. Mol. Biol.* 215:403-410). Figure 2 depicts a comparison by the PILEUP and PRETTYPLOT programs of the primary sequence of cystatin M preprotein with those of other Family 2 human cystatins; chicken cystatin was included in this alignment because its structure and function have been studied extensively. [Sequences were retrieved from the Genbank using the following accession numbers: cystatin C (A33400), cystatin D (A47142), cystatin S (S17667)] Cystatin M, like all other members of the family, shares all three conserved domains, including a conserved glycine near the N-terminal domain, Gly-36 of cystatin M preprotein (Gly-11 of mature cystatin C and Gly-9 of mature chicken cystatin). Cystatin M also contains two other structural motifs associated with cysteine proteinase inhibitory activity: the 'QXVXG' motif in the middle of the molecule, QLVAG in cystatin M (QIVAG in cystatin C, QLVSG in chicken cystatin), as well as the VPW sequence near the carboxy terminal, which is also conserved in all known cystatins. Additionally, all previously characterized members of the cystatin Family 2 contain about 120 amino acid residues with four cysteine residues near the carboxy terminal domain. These cysteine residues participate in the formation of two intrachain disulfide bridges, a characteristic structural feature of the family. Cystatin M indeed contains four cysteine residues near the carboxy terminal domain: cys-98, cys-113, cys-126 and cys-146.

The overall homology between cystatin M and other cystatin preproteins ranges from 30 to 40% for conserved amino acid residues and 25 to 33% for identical amino acids. The highest homology at the nucleotide level is 40-45% to human cystatins and 42% to chicken cystatin. A total of 30 amino acid residues are conserved and 23 residues (15%) are identical between all members of Family 2 cystatins and chicken cystatin, while 52 amino acids (30%) are identical in six out of seven proteins. Cystatin M and chicken cystatin share 48 identical amino acid residues. The overall homology of cystatin M to cystatin A and cystatin B is 25%. The homology between cystatins from different species including chicken, mouse, rat, puff adder, and is 39-48 % for conserved amino acid residues (alignment not shown in Figure 2). The novel cystatin M displays all the characterized structural features of human cystatins, and can be considered a new member of the family. Following the internationally accepted nomenclature proposal (Barrett, A.J. *et al.* (1986) *Biochem. J.* 236:312), this novel cystatin was designated cystatin M, because it was cloned from Mammary epithelial cells. Cystatin C appears the closest homolog to cystatin M. The two proteins share 33% identical and 38% conserved amino acid residues.

A Hopp and Woods hydrophilicity plot of the cystatin M protein sequence revealed the presence of a hydrophobic sequence consisting of 20 residues close to the initiation methionine which could function as a secretory signal peptide. It contains only one charged amino acid in the N-terminal region (arginine at position 3) and a cysteine at position 18. The signal peptide indicates that this protease inhibitor is synthesized as a precursor protein and probably has extracellular function(s). Cleavage of precystatin M is predicted to occur at position 21/22, with alanine at position -1 and leucine at position -3 from the cleavage site, both residues with small and uncharged side chains, resulting in mature cystatin M consisting of 127 amino acid residues. Leu-22 thus would be the putative N-terminal residue of the mature/secreted protein, although more than one native isoform differing in the length of the N-terminal sequence might exist, as has been reported for other cystatins. The predicted Mr for the precursor protein is 16,500, if the protein is not modified (*e.g.*, not glycosylated or phosphorylated) and approximately 14,300 for the putative mature protein.

To analyze the structure of the cystatin M gene, genomic DNAs were digested with restriction enzymes (EcoR1, HindIII, PvuII, NcoI) and hybridized with a cystatin cDNA probe by standard methods. A single major band hybridizing with the cystatin M cDNA probe was detected in a series of normal and tumor mammary epithelial cell lines with both EcoRI (~7.0 kb) and HindIII (~ 15.0 kb) total genomic digests. Based on these results, the cystatin M gene does not appear grossly rearranged or deleted in tumor cell lines. Down-regulation of its expression in cancerous cells is likely regulated at the transcriptional level, although extensive restriction analysis or sequencing of genomic DNA form tumor cells is required before genomic alterations or point mutations can be excluded.

EXAMPLE 2: Cell and Tissue Distribution of Cystatin M mRNA

In this example, the expression of cystatin M mRNA in various cell lines and tissues was examined by standard Northern hybridization using the cystatin M cDNA as a probe. Total cellular RNA was purified by standard guanidinium isothiocyanate and cesium chloride centrifugation. As a control for RNA loading on the hybridization filters, 36B4, a gene encoding a ribosomal protein whose expression is not affected by growth conditions or estrogen receptor expression, was used (Masiakowski, P. *et al.* (1982) *Nucl. Acids Res.* 10: 7895-7903). Densitometric scans of autoradiographs were obtained with an imaging densitometer (BioRad GS-700) using the Molecular Analyst® software.

In a first series of experiments, the expression of cystatin M mRNA was examined in exponentially growing, subconfluent normal and tumor mammary epithelial cell lines. Cell lines examined included normal human mammary epithelial cell strains (81N, 76N, and 70N) derived from reduction mammoplasty specimens, primary (21NT, 21PT), and metastatic (21MT-1, 21MT-2) tumor cell lines from the same patient and established in long-term culture, and metastatic tumor mammary epithelial cell lines MCF7, BT474, BT549, T47D, ZR-75-1, MDA-MB-157, MDA-MB-231, MDA-MB-361, MDA-MB-435, and MDA-MB-436 obtained from the American Tissue Culture Collection (ATCC/Rockville, MD). Immortal mammary epithelial cells were obtained by transfection of 76N cells with a plasmid containing the human papilloma virus (HPV)-16 E6 gene (Band, V. *et al.*, (1991) *J. Virol.* 65:6671-6676).

Representative results of the Northern blot analysis of cystatin M expression in normal and tumor human mammary epithelial cell lines are shown in Figure 3. Each lane contains 15 μg of total cellular RNA from normal and tumor cells. The blot was hybridized with a ^{32}P-labeled full-length cystatin M cDNA probe. Hybridizations were performed in formamide at 37 °C overnight. The blot was washed at 65 °C for 1 hour in 2 X SSC containing 0.1% SDS. The filter then was stripped and rehybridized to a 36B4 probe as a loading control. To summarize the overall results with mammary epithelial cell lines, cystatin M mRNA was detected in all three normal cell strains tested, 76N, 70N, and 81N, but was absent in many metastatic mammary tumor cell lines: 21MT-1, 3BT479; MCF7, ZR-75, Hs578T, T47D, BT474, BT549, MDA-MB-157, MDA-MB-361, MDA-MB-435, MDA-MB-436 (trace transcript levels were detected in MDA-MB-231). Downregulation of cystatin M mRNA in tumor cells does not seem to correlate with the estrogen receptor status. Although all normal human mammary epithelial cell strains expressed a clearly detectable cystatin M transcript, the abundance of this mRNA was well below the amount present in the overexpressing 21PT, 21NT, and 21MT-2 tumor cell lines, which were all derived from the same patient. However, the 21MT-1 cell line from the same series, which is characterized by a more metastatic phenotype, does not express cystatin M message. The cystatin M mRNA levels in human papilloma virus-immortalized normal 76N cells are comparable to the levels of its expression in the corresponding normal cells.

The expression of the cystatin M mRNA also was examined in a panel of tumor cell lines from non-mammary tissues. Cystatin M mRNA was not detected in the following tumor cell lines: PC-3 (prostate adenocarcinoma; ATCC# CRL1435); MIA Pa-CA-2 (pancreatic carcinoma; ATCC# CCL1420); HuTu 80 (duodenal adenocarcinoma); T24 (bladder transitional cell carcinoma); A549 (lung carcinoma; ATCC# CCL185); Calu-1 (lung epidermoid carcinoma); Oat 4 (lung small cell carcinoma); G-361 (malignant melanoma); SKME 30 (malignant melanoma); A2058 (malignant melanoma); SCC-25 (tongue squamous cell carcinoma); RD (rhabdomyosarcoma of pelvis); and Kaposi (Kaposi's sarcoma). Trace amount of cystatin M mRNA were detected in WiDr (ATCC# CCL218) and SW480 (ATCC# 228), both colon adenocarcinomas.

The expression of the cystatin M mRNA also was studied in normal human tissues. Representative results are shown in Figure 4, in which each lane contains approximately 2 μg of pure poly A+ RNA from the following human tissues: heart, brain placenta, lung, liver, skeletal muscle, kidney, pancreas (lanes 1,2,3,4,5,6,7, and 8, respectively). The RNAs were run on a denaturing formaldehyde/1.2% agarose gel and blotted onto a nylon membrane (Human MTN Blot, Clontech, #7760-1). The blot was hybridized to a probe corresponding to the full length cDNA of the cystatin M gene. Then, the filter was stripped and hybridized to a 36B4 probe as a loading control. The blot was washed at 65 °C for 1 hour in 2xSSC containing 0.1% SDS. Numbers in the left margin refer to the sizes of the molecular weight markers. Relatively high levels of 6A2 mRNA were present in various tissues, including placenta, lung, skeletal muscle, kidney and pancreas, although the sizes of the transcripts detected in skeletal muscle and kidney were larger. A second transcript of slightly larger size was detectable in all the above tissues. The abundance of the message was much lower in heart, while trace amounts were detectable in liver. Whether cystatin M is expressed in brain tissue cannot be determined from the blot shown in Figure 4, since the brain RNA on this blot is underloaded. Additionally, no expression of cystatin M mRNA was observed in normal breast fibroblasts (56NF cells; see Figure 3), normal foreskin fibroblasts (FS3 cells) or normal leukocytes (see Figure 3).

To determine whether there is any difference in the expression of cystatin M between exponentially growing normal mammary epithelial cells from an early passage, in early passage quiescent cells and the same cells as they became replicatively senescent, Northern hybridizations were performed with mRNA from these cells using the cystatin cDNA as a probe. Log phase, quiescent, replicative senescent and immortal cells were all derived from the 76N cell strain. The 76N cells from an early passage (passage 11) were originally plated at a very low density, harvested daily from multiple plates for isolation of RNA, measurement of cell counts and ^{3}H-thymidine incorporation assays. The plates reached full confluency after 10-11 days in culture, parallel to a plateau in cell counts. Serially passaged 76N cells from passages 15-22 were collected for RNA preparation at approximately 75 % confluency. Cells from passage 19 and after are considered

replicatively senescent, since they show a nuclear labeling index of 12 % or less and can no longer achieve a single population doubling over a time span of 2-3 weeks. The labeling index of mid-lifespan (passage 11-12) 76N cells is 22-25 % in S-phase (1 hour pulse). At passage 22, which is equivalent to 50-60 population doublings, cells can no longer be passaged and cultured for RNA isolation.

The results of this experiment are shown in Figures 5A and 5B. Figure 5A shows Northern blots containing 10 µg of total RNA from quiescent and senescent 76N cells per lane. The blot was initially hybridized to a full-length cystatin M cDNA and, subsequently, to loading control probe 36B4. Based on densitometric scans of the Northern blots in Figure 5A, levels of cystatin M expression were normalized against 36B4 control, and are presented in Figure 5B as a bar graph depicting the ratio of cystatin M/36B4. Tritium uptake for the quiescent series was calculated in cpm/cell, and is given at the bottom across the x-axis. For senescent cells, the labeling index is given as the percent (1%) of cell nuclei labeled. The results indicate that the expression of cystatin M is greatly reduced in quiescent cells as compared to replicatively senescent cells. Cystatin M mRNA level is slightly reduced in quiescent cells compared to subconfluent, actively dividing early-passage 76N cells. Cystatin M expression increases about 10-fold in senescent cells between passages 19-22 that correspond to the end of the lifespan of these cells. Accumulation of the cystatin M message is more dramatic at passage 22. The 10-fold accumulation of cystatin M mRNA in senescent cells was confirmed with a serially passaged 76N senescent cell series from an independent experiment.

EXAMPLE 3: Expression of Recombinant Cystatin M

In this example, cystatin M was expressed as a recombinant glutathione-S-transferase (GST) fusion protein in *Escherichia coli* and the fusion protein was isolated and characterized. The cystatin M open reading frame cDNA sequence encoding the putative mature protein (Leu22-Met149) was amplified by PCR using a sense primer having the nucleotide sequence: 5'-GGAATTCTGCCACGAGATGCCCGGGC-3' (SEQ ID NO: 3), and an antisense primer having the nucleotide sequence: 5'-CCCTCGAATTCTTATCACAT CTGCAC-3' (SEQ ID NO: 4). This pair of gene-specific synthetic oligonucleotides correspond to the sequences on the sense strand upstream of the ATG start site and to the antisense strand downstream to the stop codon. The N-terminal amino acid consists of the residue Leu-22. Thus, the amplified region of the cDNA sequence does not contain the hydrophobic signal peptide of cystatin M. These 26-mer oligonucleotide primers were designed to create EcoRI overhangs on the resultant PCR product, thereby allowing for cloning of the PCR product into an EcoRI site. The amplification included two initial cycles at low stringency (42 °C) and thirty-eight cycles at higher stringency (60 °C). The amplified product was sequenced to ensure that it contained no PCR-induced mutations. The PCR product was then digested with *Eco*RI and ligated

into the pGEX-2T vector that was *Eco*RI-digested and linearized (Pharmacia Biotech Inc.) (Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40). This resulted in the expression plasmid pGEX-2T/cystatin M, encoding a fusion protein comprising, from the N-terminus to the C-terminus: GST-a thrombin cleavage site-cystatin M.

E. coli XL-1 Blue bacteria (Stratagene, La Jolla, CA) were transformed with either the parental vector (pGEX-2T) or the recombinant vector (pGEX-2T/cystatin M) and propagated in Luria Broth (LB) (Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd. ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor, NY, 1989) in the presence of 100 μg/ml ampicillin for selection of the cells transformed with the vector. The expression of recombinant fusion protein was induced in exponentially growing bacteria (A550=0.8-1.0) with IPTG, at a final concentration of 0.2 mM, and the bacteria were incubated for 1.5 hour at 37 °C with vigorous agitation. The bacteria were harvested by centrifugation, washed twice with MTPBS and resuspended in lysis buffer, MTPBS (150 mM NaCl, 16 mM Na_2HPO_4, 4mM NaH_2PO_4 pH 7.4) containing 1 mM DTT, 1 mM PMSF, and 2 % Triton X-100. Cells were lysed on ice by mild sonication and the suspension was centrifuged at 14,500xg for 15 min to remove unlysed cells. All subsequent purification steps were carried out at 4 °C.

The recombinant fusion protein was purified from crude bacterial lysates by affinity chromatography on glutathione agarose resin (Sigma Chemical Co., St. Louis, MO). The clear bacterial lysate was collected and bound by glutathione agarose beads. The glutathione column was washed with MTPBS containing 350 mM NaCl and the fusion protein was eluted with 50 mM Tris.HCl pH 8.0 containing 5 mM reduced glutathione. Purified rGST-Cystatin M was dialyzed against MTPBS containing 10% glycerol, sterilized by filtration through 0.22 μm filters (Costar, Cambridge, MA), and stored at -20 °C. The concentration of the purified protein was determined by Bradford assay (BioRad) using γ-globulin as a standard. The fusion protein (rGST-cystatin M) was sufficiently soluble to be purified from crude bacterial lysates by affinity chromatography on a glutathione affinity column under non-denaturing conditions with estimated yield of approximately 3-5 mg per liter of bacterial culture.

The protein composition of the lysate was analyzed by SDS-PAGE. The reagents for protein SDS-PAGE analysis and protein concentration determination were purchased from BioRad, Hercules, CA. The purity of rGST-Cystatin M was assessed by Coomassie and Silver (Silver Stain Plus, BioRad) staining of 15% acrylamide SDS gels. The results of this SDS-PAGE analysis are shown in Figure 6, lanes 1-4. Lane 1 shows lysates from cells transformed with the parental pGEX-2T vector. Lane 2 shows the material from the pGEX-2T lysate that bound to the glutathione-agarose resin. Lane 3 shows lysates from cells transformed with the pGEX-2T/cystatin M expression vector. Lane 4 shows the material from the pGEX-2T/cystatin M lysate that bound to the glutathione-agarose resin. The results demonstrate that lysates of bacteria transformed with pGEX-2T/cystatin M

contained the rGST-cystatin M fusion protein, which was resolved as a single band of Mr 41 kDa on an SDS gel under reducing conditions. This band was not present in the control extracts from bacteria transformed with the pGEX-2T parental vector, which contained only the rGST protein of 26 kDa.

The rGST carrier was completely cleaved from the purified fusion protein by proteolytic digestion at the thrombin site located between the GST and cystatin M portions of the fusion protein. Thrombin treatment was carried out at room temperature in the presence of 150 mM NaCl, 2.1 mM $CaCl_2$, with 1 µg/µl fusion protein and 3.2 NIH thrombin units/ml for about 2 hours. The reaction was stopped with 0.1 mM EGTA and cleavage was monitored by SDS-PAGE. The rCystatin M was further purified by absorption of rGST and any traces of uncleaved rGST-cystatin M on immobilized glutathione. The thrombin-cleaved material was analyzed by SDS-PAGE, the results of which are shown in Figure 6, lanes 5-8, which shown the cleavage products after thrombin treatment for 0, 2, 30 or 90 minutes, respective. The rCystatin M cleaved from the fusion protein was resolved on a reducing SDS gel as a single band with a M(r) of approximately 14 kDa.

EXAMPLE 4: Preparation and Use of an Anti-Cystatin M Antibody

In this example, a polyclonal antisera was raised against the recombinant GST-cystatin M fusion protein described in the previous example. The purified fusion protein was used to immunize New Zealand white rabbits 3-9 months old. Antiserum was recovered from the immunized animals and affinity-purified, first on a GST-glutathione agarose column (to remove antibodies specific for the GST portion of the fusion protein) and then on a GST/Cystatin M-glutathione agarose column (to select for antibodies specific for the cystatin M portion of the fusion protein) (as described in Krek, W. *et al.* (1993) *Science* 262:1557-1560). The purified antibody was dialyzed against PBS containing 50% glycerol, and 0.02% NaN_3 was added, and stored at 4 °C.

The anti-cystatin M antisera was used in immunoprecipitation and Western blot (immunoblot) experiments. For the preparation of whole cell lysates, cells were washed with PBS and resuspended (8-10 x 10^6 cells/ml) in lysis buffer (50 mM Tris pH 8.0, containing 120 mM NaCl, 0.5% Nonidet P-40, 5 µg/ml leupeptin, Na-ortho-vanidate, and 100 mM NaF). Lysed cells were rocked for 30 min, centrifuged at 14,000 x rpm for 15 min, and the supernatants were assayed immediately. All steps were performed at 4 °C. For immunoprecipitation, 250 µl of fresh total cell lysate were diluted 1:1 with 20 mM Tris pH 8.0, containing 100 mM NaCl, 1 mM EDTA, and 0.5 % Nonidet P-40, preimmune serum and affinity-purified antiserum were added, respectively, to each 250 µl sample at a 1:250 dilution, and the samples were incubated with mild agitation for one hour. The immunoprecipitated proteins were then bound to Protein A Sepharose beads for 30 min, solubilized in sample buffer at 90 °C and analyzed by SDS-PAGE. Native cystatin M was

immunoprecipitated from 76N and 21PT cell culture supernatants using affinity-purified antiserum, accordingly.

Alternatively, proteins from cell culture were precipitated with trichloroacetic acid, resuspended in lysis buffer and analyzed by Western blotting. Protease inhibitors were added to the supernatants immediately after collection to prevent proteolytic degradation of cystatin M and were stored at -70 °C. For immunoblot detection, proteins separated by SDS-PAGE were transferred to PDVF membrane (0.2 micron, BioRad) and reacted with polyclonal antiserum and preimmune serum. Anti-rabbit Ig, horseraddish peroxidase linked, whole antibody was used as secondary antibody (1:2000) and immunoreactive proteins were detected with the ECI system (enhanced chemiluminescence, Amersham). Transfer and quantitation of proteins were assessed by staining with 0.1 %w/v amidoblack in 25 % isopropanol and 10 % acetic acid and destained with 50 % methanol and 7.5 % acetic acid in H_2O.

Representative Western blot and immunoprecipitation results are shown in Figures 7A and 7B, respectively. Figure 7A illustrates Western blot detection of cystatin M protein levels in either the lysate (L) or supernatant (S) from a normal human mammary epithelial cell line (70N), a primary mammary epithelial tumor cell line (21PT) or a malignant mammary epithelial tumor cell line (MDA435). The Western blot results indicate that cystatin M protein expression in normal and tumor mammary cell lines parallels the mRNA expression patterns described in Example 2, namely that cystatin M protein is detectable in the normal mammary epithelial cell line and the primary mammary epithelial tumor cell line but is not detectable in the metastatic tumor cell line. Figure 7B illustrates representative immunoprecipitates with preimmune serum (lanes 1 and 3) and with affinity-purified polyclonal antibody raised against rGST-cystatin M (lanes 2 and 4) from culture supernatants of 21PT cells (lanes 1 and 2) or MDA435 cells (lanes 3 and 4). The results of the immunoprecipitation experiments indicate the presence of a native immunoreactive protein in the supernatant of 21PT cells, but not MDA435 cells, whose size is consistent with the predicted size of cystatin M. mature protein, demonstrating that 21PT cells express and secrete a native protein corresponding in size and immunoreactivity to cystatin M.

EXAMPLE 5: Cysteine Proteinase Inhibitor Activity of Cystatin M

In this example, the ability of cystatin M to inhibit the cysteine proteinase activity of papain was examined. Papain isolated from papaya latex was purchased from Boehringer Mannheim. Papain activity was assayed at room temperature in 125 mM phosphate buffer, pH 6.8 containing 4 mM DTT, 1 mM EDTA, and 0.05% Brij 35 using the fluorogenic synthetic substrate Z-Phe-Arg-MCA (purchased from Sigma Chemical Co., St. Louis, MO). To determine the inhibitory effect of cystatin M on papain activity, papain solutions (5-100 pM or 10 nM) were preincubated with increasing amounts of rGST-cystatin M (0-5 nM or 0-1 μM, prepared as described in Example 3), in a total volume of

50 µl; for 5 min at room temperature, then added to 2 ml assay buffer containing 2.5-150 µM Z-Phe-Arg-MCA substrate. The reaction mixture was stirred during the assay. The initial reaction rates were monitored for 1-2 min by the increase in the intensity of relative fluorescence with a fluorimeter (Model SFM25, Kontron Instruments). The excitation and emission wavelengths were 380nm and 440nm, respectively. The amount of 7-amino-4-methylcoumarin liberated from the synthetic substrate was determined from a standard curve. The concentration of active cysteine proteinase in the papain solution was determined by active-site titration with E-64 L-3-carboxy-*trans*-2,3-epoxy-propionyl-leucylamido-(4-guanidino)butane (Barrett, A.J. and Kirschke, H. (1981) *Meth. Enzymol.* 80:535-561). Under the conditions used, the self-hydrolysis of the Z-Phe-Arg-MCA substrate was negligible for the applied reaction times. As a negative control, the inhibitory effects of GST and BSA were tested under the same assay conditions.

The results, illustrated graphically in Figure 8A, demonstrate that GST-cystatin M displays inhibitory activity against the prototype cysteine proteinase papain using Z-Phe-Arg-MCA synthetic substrate, since the amount of liberated fluorescent material is significantly greater in the absence of any inhibitors ("no I" in Figure 8A) than in the presence of 3 nM cystatin M ("[I]=3nM" in Figure 8A). The results of inhibition assays with different concentrations of substrate and cystatin M fusion protein are plotted as a Dixon plot in Figure 8B. [S] depicts the concentration of the substrate in µM. v, is the initial reaction rate and was expressed by the increase of relative fluorescence intensity at 440 nm per im, with an excitation wavelength at 380 nm. To obtain inhibition constants (Ki), the reciprocal of v (1/v) was plotted against the concentration of rGST-cystatin M expressed in nM [FP] at different substrate concentrations [S], according to the method of Dixon. The Ki value corresponds to the inhibitor concentration at which the three lines intersect, with computer-aided linear regression, using Cricket graph. Linear regression coefficients were greater than 0.980. All determinations were based on assays with less than 1% substrate self-hydrolysis at the applied substrate concentration. A Ki value of 0.5 nM was estimated from Dixon plots of continuous and stopped flow assays at 100 pM papain, suggestive of tight binding of native cystatin M to papain.

EXAMPLE 6: Glycosylation of Cystatin M

Immunoprecipitation of native cystatin M from 21PT cells using an anti-cystatin M antibody detected a protein of approximately 14.5 kDa, consistent with the predicted size for the cystatin M gene product, and a second immunoreactive protein of approximately 20-22 kDa. To determine whether this 20-22 kDa protein represents a glycosylated form of native cystatin M, the protein was treated with N-Glycosidase F. Cystatin M protein immunoprecipitated from 21PT cell culture supernatant was treated with N-Glycosidase F for 24 hours at 37° C. The protein was then analyzed by Western blot, the results of which are shown in Figure 9. Lane 1 represents native cystatin M not treated with N-Glycosidase

F. Lane 2 represents native cystatin M treated with N-Glycosidase F. Lane 3 represents cystatin M cleaved from the recombinant GST-cystatin M (rGST-cystatin M) fusion protein by thrombin. For Western detection, the proteins were immunoblotted with anti-rGST-cystatin M serum. Bound antibody was detected by ECL (Amersham). The molecular weights of the immunoreactive proteins were estimated based on their mobility relative to prestained molecular weight markers shown in kilobases on the left in Figure 9. Cystatin M cleaved from rGST-cystatin M (lane 3) migrated with the expected apparent size of 14.5 kDa. The untreated native cystatin M (lane 1) comprises both the 14.5 kDa form and the 20-22 kDa form. Treatment of the native cystatin M with N-Glycosidase F (lane 2) completely abolished the 20-22 kDa form, indicating that this form represents an N-glycosylated form of cystatin M.

The N-glycosylated form accounts for about 30-40% of total native cystatin M protein in 21PT cells. A potential site for N-linked glycosylation of cystatin M is Asn^{137}, near the C-terminus and in close proximity to the conserved Val^{133}-Pro-Trp^{135} motif. Asn^{137} is located between the cysteine residues that form the disulfide bridge, which is important in maintaining the conformation required for inhibitory activity of cystatins. The increase in size of glycosylated cystatin M by 6 kDa could be accounted for by two carbohydrate moieties, although the presence of charged sugars like sialic acid would change the net charge of the protein and thus modify its electrophoretic mobility.

EXAMPLE 7: Recombinant Expression of Cystatin M in a Metastatic Breast Cancer Cell Line

The mammary epithelial tumor cell line MDA-MB-435 (Price *et al.* (1990) *Cancer Res.* 50:717), which does not express endogenous cystatin M mRNA transcript, was used as the host cell to create a series of stable cystatin M transfectant clones. DNA encoding cystatin M was introduced into the expression vector pCMVneo and the resultant cystatin M expression vector (pCMVneo/cystatin M) was transfected into MDA-MB-435 cells. Stable clones were selected with G418. The clones express varying levels of cystatin M transcript and protein (similar, or 2-30 fold higher than endogenous cystatin M in 21PT cells, the cell line from which the cystatin M cDNA was cloned). The protein produced by the transfectants is secreted and glycosylated, similar to the endogenous protein in 21PT cells. The recombinant protein can be isolated from the culture supernatant by standard methods.

Stable cystatin M transfectants of metastatic breast tumor cells (*e.g.*, the above-described MDA-MB-435 transfectants) can be used to evaluate phenotypic changes induced by cystatin M, such as the invasive potential of the cells, migration of the cells, growth rate in culture and tumor formation in nude mice. The parental cell line (*e.g.*, MDA-MB-435 cells) or parental vector transfectants (*e.g.*, MDA-MB-435 cells transfected with pCMVneo) can be used as controls. Preliminary experiments showed that the growth rates of most of the pCMVneo/cystatin M MDA-MB-435 transfectants in culture were significantly inhibited, as compared to pCMVneo and parental controls. To evaluate tumor formation *in vivo*, the MDA-MB-435 transfectants can be injected into nude mice. The parental cell line is known to form tumors at the site of orthotopic injection and metastasizes in nude mice (described further in Price *et al.* (1990) *Cancer Res.* 50:717). At 10-weeks post-inoculation, the mice can be sacrificed and their tumors excised and weighed to determine the effect of cystatin M expression on tumor progression and metastasis. A suitable *in vitro* assay is tumor cell invasion through reconstituted basement membrane matrix (*e.g.*, Matrigel) as described in Hendrix *et al.* (1987) *Cancer Letters* 38:137. The invasive ability of cystatin M-transfected MDA-MB-435 cells can be compared to untransfected MDA-MB-435 cells to determine the effect of cystatin M expression on tumor invasiveness.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: DANA-FARBER CANCER INSTITUTE
(B) STREET: 44 BINNEY STREET
(C) CITY: BOSTON
(D) STATE: MASSACHUSETTS
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 02115
(G) TELEPHONE: (617) 632-4016
(H) TELEFAX: (617) 632-4012

(ii) TITLE OF INVENTION: Cystatin M, A Novel Cysteine Proteinase Inhibitor

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: LAHIVE & COCKFIELD
(B) STREET: 60 State Street, suite 510
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02109-1875

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/546,000
(B) FILING DATE: 20-OCT-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: DeConti, Giulio A. ,Jr.
(B) REGISTRATION NUMBER: 31,503
(C) REFERENCE/DOCKET NUMBER: DFN-004PC

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617)227-7400
(B) TELEFAX: (617)227-5941

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 598 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 24..470

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GAGCTCCGAC GGCACTGACG GCC ATG GCG CGT TCG AAC CTC CCG CTG GCG          50
                          Met Ala Arg Ser Asn Leu Pro Leu Ala
                            1               5

CTG GGC CTG GCC CTG GTC GCA TTC TGC CTC CTG GCG CTG CCA CGC GAT        98
Leu Gly Leu Ala Leu Val Ala Phe Cys Leu Leu Ala Leu Pro Arg Asp
 10                  15                  20                  25

GCC CGG GCC CGG CCG CAG GAG CGC ATG GTC GGA GAA CTC CGG GAC CTG       146
Ala Arg Ala Arg Pro Gln Glu Arg Met Val Gly Glu Leu Arg Asp Leu
                 30                  35                  40

TCG CCC GAC GAC CCG CAG GTG CAG AAG GCG GCG CAG GCG GCC GTG GCC       194
Ser Pro Asp Asp Pro Gln Val Gln Lys Ala Ala Gln Ala Ala Val Ala
             45                  50                  55

AGC TAC AAC ATG GGC AGC AAC AGC ATC TAC TAC TTC CGA GAC ACG CAC       242
Ser Tyr Asn Met Gly Ser Asn Ser Ile Tyr Tyr Phe Arg Asp Thr His
         60                  65                  70

ATC ATC AAG GCG CAG AGC CAG CTG GTG GCC GGC ATC AAG TAC TTC CTG       290
Ile Ile Lys Ala Gln Ser Gln Leu Val Ala Gly Ile Lys Tyr Phe Leu
     75                  80                  85

ACG ATG GAG ATG GGG AGC ACA GAC TGC CGC AAG ACC AGG GTC ACT GGA       338
Thr Met Glu Met Gly Ser Thr Asp Cys Arg Lys Thr Arg Val Thr Gly
 90                  95                 100                 105

GAC CAC GTC GAC CTC ACC ACT TGC CCC CTG GCA GCA GGG GCG CAG CAG       386
Asp His Val Asp Leu Thr Thr Cys Pro Leu Ala Ala Gly Ala Gln Gln
                110                 115                 120

GAG AAG CTG CGC TGT GAC TTT GAG GTC CTT GTG GTT CCC TGG CAG AAC       434
Glu Lys Leu Arg Cys Asp Phe Glu Val Leu Val Val Pro Trp Gln Asn
            125                 130                 135

TCC TCT CAG CTC CTA AAG CAC AAC TGT GTG CAG ATG T GATAAGTCCC          481
Ser Ser Gln Leu Leu Lys His Asn Cys Val Gln Met
        140                 145

CGAGGGCGAA GGCCATTGGG TTTGGGGCCA TGGTGGAGGG CACTTCAGGT CCGTGGGCCG     541

TATCTGTCAC AATAAATGGC CAGTGCTGCT TCTTGCAAAA AAAAAAAAAA AAAAAAA        598
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 149 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala Arg Ser Asn Leu Pro Leu Ala Leu Gly Leu Ala Leu Val Ala
 1               5                  10                  15

Phe Cys Leu Leu Ala Leu Pro Arg Asp Ala Arg Ala Arg Pro Gln Glu
            20                  25                  30

Arg Met Val Gly Glu Leu Arg Asp Leu Ser Pro Asp Asp Pro Gln Val
        35                  40                  45

Gln Lys Ala Ala Gln Ala Ala Val Ala Ser Tyr Asn Met Gly Ser Asn
    50                  55                  60

Ser Ile Tyr Tyr Phe Arg Asp Thr His Ile Ile Lys Ala Gln Ser Gln
65                  70                  75                  80

Leu Val Ala Gly Ile Lys Tyr Phe Leu Thr Met Glu Met Gly Ser Thr
                85                  90                  95

Asp Cys Arg Lys Thr Arg Val Thr Gly Asp His Val Asp Leu Thr Thr
            100                 105                 110

Cys Pro Leu Ala Ala Gly Ala Gln Gln Glu Lys Leu Arg Cys Asp Phe
        115                 120                 125

Glu Val Leu Val Val Pro Trp Gln Asn Ser Ser Gln Leu Leu Lys His
    130                 135                 140

Asn Cys Val Gln Met
145
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGAATTCTGC CACGAGATGC CCGGGC 26

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 bases

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CCCTCGAATT CTTATCACAT CTGCAC 26

CLAIMS

1. An isolated nucleic acid molecule comprising a nucleotide sequence encoding cystatin M or a biologically active portion thereof.

2. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2 and inhibits the activity of papain *in vitro*.

3. The isolated nucleic acid molecule of claim 2, wherein the protein comprises an amino acid sequence at least 70 % homologous to the amino acid sequence of SEQ ID NO: 2.

4. The isolated nucleic acid molecule of claim 2, wherein the protein comprises an amino acid sequence at least 80 % homologous to the amino acid sequence of SEQ ID NO: 2.

5. The isolated nucleic acid molecule of claim 2, wherein the protein comprises an amino acid sequence at least 90 % homologous to the amino acid sequence of SEQ ID NO: 2.

6. An isolated nucleic acid molecule at least 15 nucleotides in length which hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1.

7. The isolated nucleic acid molecule of claim 6 which comprises a naturally-occurring nucleotide sequence.

8. The isolated nucleic acid molecule of claim 6 which encodes human cystatin M.

9. The isolated nucleic acid molecule of claim 6 which encodes mouse cystatin M.

10. An isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1.

11. The isolated nucleic acid molecule of claim 8, comprising the coding region of the nucleotide sequence of SEQ ID NO: 1.

12. An isolated nucleic acid molecule encoding the amino acid sequence of SEQ ID NO: 2.

13. An isolated nucleic acid molecule encoding a cystatin M fusion protein.

14. An isolated nucleic acid molecule which is antisense to the nucleic acid molecule of claim 1.

15. An isolated nucleic acid molecule which is antisense to the coding strand of the nucleic acid molecule of claim 10 comprising the nucleotide sequence of SEQ ID NO: 1.

16. The isolated nucleic acid molecule of claim 15 which is antisense to a coding region of the coding strand of the nucleotide sequence of SEQ ID NO: 1.

17. The isolated nucleic acid molecule of claim 15 which is antisense to a noncoding region of the coding strand of the nucleotide sequence of SEQ ID NO: 1.

18. A vector comprising a nucleotide sequence encoding cystatin M.

19. The vector of claim 18, which is a recombinant expression vector.

20. The vector of claim 19, which encodes a protein comprising the amino acid sequence of SEQ ID NO: 2.

21. The vector of claim 19, which comprises the coding region of the nucleotide sequence of SEQ ID NO: 1.

22. A host cell containing the vector of claim 18.

23. A host cell containing the recombinant expression vector of claim 19.

24. A method for producing cystatin M comprising culturing the host cell of claim 23 in a suitable medium until cystatin M is produced.

25. The method of claim 24, further comprising isolating cystatin M from the medium or the host cell.

26. An isolated cystatin M protein or a biologically active portion thereof.

27. The isolated cystatin M protein of claim 26, which is glycosylated.

28. An isolated protein which comprises an amino acid sequence at least 60 % homologous to the amino acid sequence of SEQ ID NO: 2 and inhibits the activity of papain *in vitro*.

29. An isolated protein comprising amino acids 1-149 of SEQ ID NO: 2.

30. The isolated protein of claim 29, comprising about amino acids 22-149 of SEQ ID NO: 2.

31. A pharmaceutical composition comprising the protein of claim 30 and a pharmaceutically acceptable carrier.

32. A fusion protein comprising a cystatin M polypeptide operatively linked to a non-cystatin M polypeptide.

33. An antigenic peptide of cystatin M comprising at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO: 2, the peptide comprising an epitope of cystatin M such that an antibody raised against the peptide forms a specific immune complex with cystatin M.

34. An antibody that specifically binds cystatin M.

35. The antibody of claim 34, which is monoclonal.

36. The antibody of claim 35, which is coupled to a detectable substance.

37. A pharmaceutical composition comprising the antibody of claim 34 and a pharmaceutically acceptable carrier.

38. A nonhuman transgenic animal which contains cells carrying a transgene encoding cystatin M.

39. A nonhuman homologous recombinant animal which contains cells having an altered cystatin M gene.

40. A method for detecting the presence of cystatin M in a biological sample comprising contacting a biological sample with an agent capable of detecting cystatin M protein or mRNA such that the presence of cystatin M is detected in the biological sample.

41. The method of claim 40, wherein the agent is a labeled or labelable nucleic acid probe capable of hybridizing to cystatin M mRNA.

42. The method of claim 40, wherein the agent is a labeled or labelable antibody capable of specifically binding to cystatin M protein.

43. The method of claim 40, wherein the biological sample is a tumor sample.

44. A method for diagnosis of a subject with a tumor comprising:

contacting a tumor sample from the subject with an agent capable of detecting cystatin M protein or mRNA;

determining the amount of cystatin M protein or mRNA expressed in the tumor sample;

comparing the amount of cystatin M protein or mRNA expressed in the tumor sample to a control sample; and

forming a diagnosis based on the amount of cystatin M protein or mRNA expressed in the tumor sample as compared to the control sample.

45. The method of claim 44, wherein the tumor sample is a mammary tumor sample.

46. A kit for detecting the presence of cystatin M in a biological sample comprising a labeled or labelable agent capable of detecting cystatin M protein or mRNA in a biological sample; means for determining the amount of cystatin M in the sample; and means for comparing the amount of cystatin M in the sample with a standard.

47. The kit of claim 46, wherein the agent is a nucleic acid probe capable of hybridizing to cystatin M mRNA.

48. The kit of claim 46, wherein the agent is an antibody capable of specifically binding to cystatin M protein.

49. A method comprising contacting a cell with an agent that modulates cystatin M cysteine proteinase inhibitory activity associated with the cell.

50. The method of claim 49, wherein the agent stimulates the cystatin M cysteine proteinase inhibitory activity associated with the cell.

51. The method of claim 50, wherein the agent is an active cystatin M protein.

52. The method of claim 50, wherein the agent is a nucleic acid encoding cystatin M that has been introduced into the cell.

53. The method of claim 49, wherein the agent inhibits the cystatin M cysteine proteinase inhibitory activity associated with the cell.

54. The method of claim 53, wherein the agent is an antisense cystatin M nucleic acid molecule.

55. The method of claim 53, wherein the agent is an antibody that specifically binds to cystatin M.

56. The method of claim 49, wherein the cell is present within a subject and the agent is administered to the subject.

57. A method for inhibiting development or progression of a metastatic phenotype in a tumor cell comprising contacting the tumor cell with an agent which elevates the amount of cystatin M in or around the tumor cell.

58. The method of claim 57, wherein the agent is cystatin M.

59. The method of claim 57, wherein the agent is a nucleic acid encoding cystatin M that has been introduced into the tumor cell.

60. The method of claim 57, wherein the tumor cell is a mammary tumor cell.

61. A method for identifying a modulator of the cysteine proteinase inhibitory activity of cystatin M, comprising

incubating cystatin M, a cysteine proteinase, a substrate for the cysteine proteinase and a test substance under conditions suitable for the cysteine proteinase to cleave the substrate;

measuring the cleavage of the substrate;

comparing the amount of cleavage of the substrate in the presence of the test substance to the amount of cleavage of the substrate in the absence of the test substance; and

identifying the test substance as a modulator of the cysteine proteinase inhibitory activity of cystatin M.

62. A method for identifying a modulator of cystatin M expression, comprising

contacting a cell with a test substance;

determining the level of expression of cystatin M mRNA or protein in the cell;

comparing the level of expression of cystatin M mRNA or protein in the cell in the presence of the test substance to level of expression of cystatin M mRNA or protein in the cell in the absence of the test substance; and

identifying the test substance as a modulator of cystatin M expression.

```
  1 GAGCTCCGACGGCACTGACGGCC                                          23

 24 ATG GCG CGT TCG AAC CTC CCG CTG GCG CTG GGC CTG GCC              62
  1 Met Ala Arg Ser Asn Leu Pro Leu Ala Leu Gly Leu Ala              13

 63 CTG GTC GCA TTC TGC CTC CTG GCG CTG CCA CGC GAT GCC 101
 14 Leu Val Ala Phe Cys Leu Leu Ala Leu Pro Arg Asp Ala  26

102 CGG GCC CGG CCG CAG GAG CGC ATG GTC GGA GAA CTC CGG 140
 27 Arg Ala Arg Pro Gln Glu Arg Met Val Gly Glu Leu Arg  39

141 GAC CTG TCG CCC GAC GAC CCG CAG GTG CAG AAG GCG GCG 179
 40 Asp Leu Ser Pro Asp Asp Pro Gln Val Gln Lys Ala Ala  52

180 CAG GCG GCC GTG GCC AGC TAC AAC ATG GGC AGC AAC AGC 218
 53 Gln Ala Ala Val Ala Ser Tyr Asn Met Gly Ser Asn Ser  65

219 ATC TAC TAC TTC CGA GAC ACG CAC ATC ATC AAG GCG CAG 257
 66 Ile Tyr Tyr Phe Arg Asp Thr His Ile Ile Lys Ala Gln  78

258 AGC CAG CTG GTG GCC GGC ATC AAG TAC TTC CTG ACG ATG 296
 79 Ser Gln Leu Val Ala Gly Ile Lys Tyr Phe Leu Thr Met  91

297 GAG ATG GGG AGC ACA GAC TGC CGC AAG ACC AGG GTC ACT 335
 92 Glu Met Gly Ser Thr Asp Cys Arg Lys Thr Arg Val Thr 104

336 GGA GAC CAC GTC GAC CTC ACC ACT TGC CCC CTG GCA GCA 374
105 Gly Asp His Val Asp Leu Thr Thr Cys Pro Leu Ala Ala 117

375 GGG GCG CAG CAG GAG AAG CTG CGC TGT GAC TTT GAG GTC 413
118 Gly Ala Gln Gln Glu Lys Leu Arg Cys Asp Phe Glu Val 130

414 CTT GTG GTT CCC TGG CAG AAC TCC TCT CAG CTC CTA AAG 452
131 Leu Val Val Pro Trp Gln Asn Ser Ser Gln Leu Leu Lys 143

453 CAC AAC TGT GTG CAG ATG                                        470
144 His Asn Cys Val Gln Met                                        149

471 TGATAAGTCCCCGAGGGCGAAGGCCATTGGGTTTGGGGCCATGGTGGAGGG 521
522 CACTTCAGGTCCGTGGGCCGTATCTGTCACAATAAATGGCCAGTGCTGCTT 572
573 CTTGCAAAAAAAAAAAAAAAAAAA                                     598
```

FIGURE 1

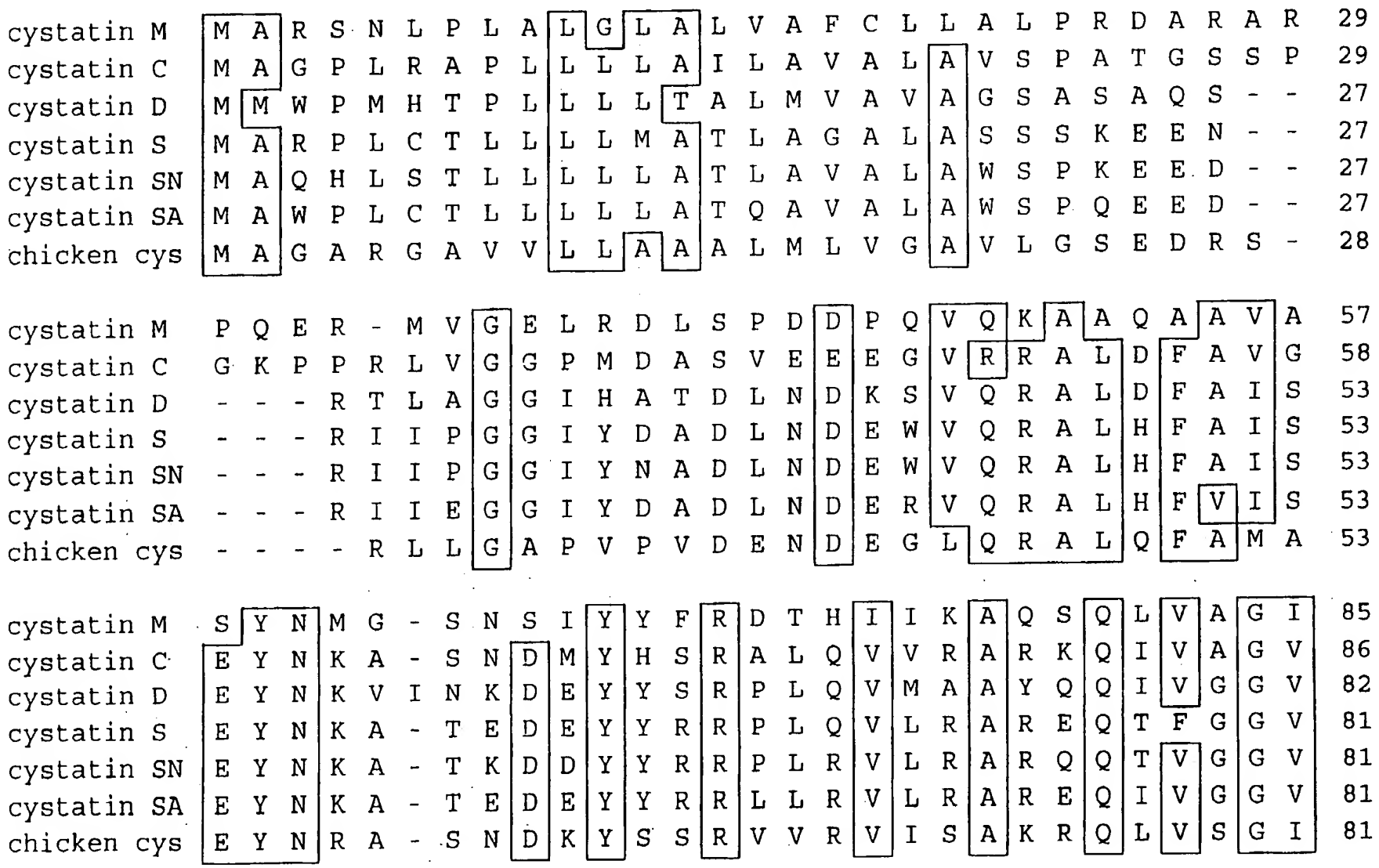

cystatin M M A R S N L P L A L G L A L V A F C L L A L P R D A R A R 29
cystatin C M A G P L R A P L L L L A I L A V A L A V S P A T G S S P 29
cystatin D M M W P M H T P L L L L T A L M V A V A G S A S A Q S - - 27
cystatin S M A R P L C T L L L L M A T L A G A L A S S S K E E N - - 27
cystatin SN M A Q H L S T L L L L L A T L A V A L A W S P K E E D - - 27
cystatin SA M A W P L C T L L L L L A T Q A V A L A W S P Q E E D - - 27
chicken cys M A G A R G A V V L L A A A L M L V G A V L G S E D R S - 28
cystatin M P Q E R - M V G E L R D L S P D D P Q V Q K A A Q A A V A 57
cystatin C G K P P R L V G G P M D A S V E E E G V R R A L D F A V G 58
cystatin D - - - R T L A G G I H A T D L N D K S V Q R A L D F A I S 53
cystatin S - - - R I I P G G I Y D A D L N D E W V Q R A L H F A I S 53
cystatin SN - - - R I I P G G I Y N A D L N D E W V Q R A L H F A I S 53
cystatin SA - - - R I I E G G I Y D A D L N D E R V Q R A L H F V I S 53
chicken cys - - - - R L L G A P V P V D E N D E G L Q R A L Q F A M A 53
cystatin M S Y N M G - S N S I Y Y F R D T H I I K A Q S Q L V A G I 85
cystatin C E Y N K A - S N D M Y H S R A L Q V V R A R K Q I V A G V 86
cystatin D E Y N K V I N K D E Y Y S R P L Q V M A A Y Q Q I V G G V 82
cystatin S E Y N K A - T E D E Y Y R R P L Q V L R A R E Q T F G G V 81
cystatin SN E Y N K A - T K D D Y Y R R P L R V L R A R Q Q T V G G V 81
cystatin SA E Y N K A - T E D E Y Y R R L L R V L R A R E Q I V G G V 81
chicken cys E Y N R A - S N D K Y S S R V V R V I S A K R Q L V S G I 81

FIGURE 2A

```
cystatin M   K Y F L T M E M G S T D C R K T R V T G D H V D L T T C P 114
cystatin C   N Y F L D V E L G R T T C T K T - - - - - Q P N L D N C P 110
cystatin D   N Y F F N V K F G R T T C T K S - - - - - Q P N L D N C P 106
cystatin S   N Y F F D V E V G R T I C T K S - - - - - Q P N L D T C A 105
cystatin SN  N Y F F D I E V G R T I C T K S - - - - - Q P N L D T C A 105
cystatin SA  N Y F F D I E V G R T I C T K S - - - - - Q P N L D T C A 105
chicken cys  K Y I L Q V E I G R T T C P K S - - - - - S G D L Q S C E 105

cystatin M   L A A G A Q - Q E K L R C D F E V L V V P W Q N S S Q L L 142
cystatin C   F H D Q P H L K R K A F C S F Q I Y A V P W Q G T M T L S 139
cystatin D   F H D Q P K L K E E E F C S F Q I N E V P W E D K I S I L 135
cystatin S   F H E Q P E L Q K K Q L C S F E I Y E V P W E D R M S L V 134
cystatin SN  F H E Q P E L Q K K Q L C S F E I Y E V P W E N R R S L V 134
cystatin SA  F H E Q P E L Q K K Q L C S F Q I Y E V P W E D R M S L V 134
chicken cys  F H D E P E M A K Y T T C T F V V Y S I P W L N Q I K L L 134

cystatin M   K H N C V Q M   149
cystatin C   K S T C Q D A   146
cystatin D   N Y K C R K V   142
cystatin S   N S R C Q E A   141
cystatin SN  K S R C Q E S   141
cystatin SA  N S R C Q E A   141
chicken cys  E S K C Q - -   139
```

FIGURE 2B

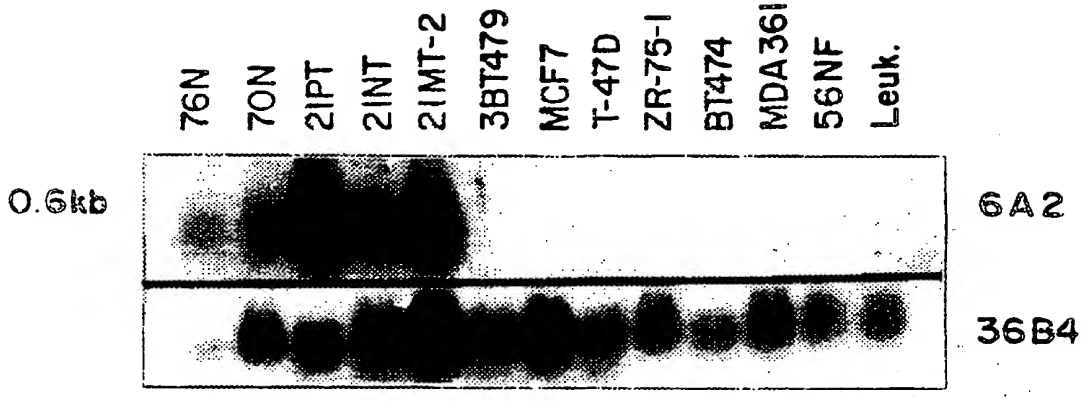

76N
70N
21PT
21NT
21MT-2
3BT479
MCF7
T-47D
ZR-75-1
BT474
MDA361
56NF
Leuk.
0.6kb
6A2
36B4

FIG.3

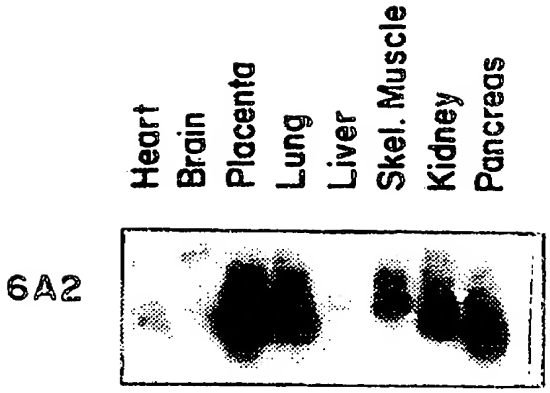

Heart
Brain
Placenta
Lung
Liver
Skel. Muscle
Kidney
Pancreas
6A2

FIG.4

FIG.5A

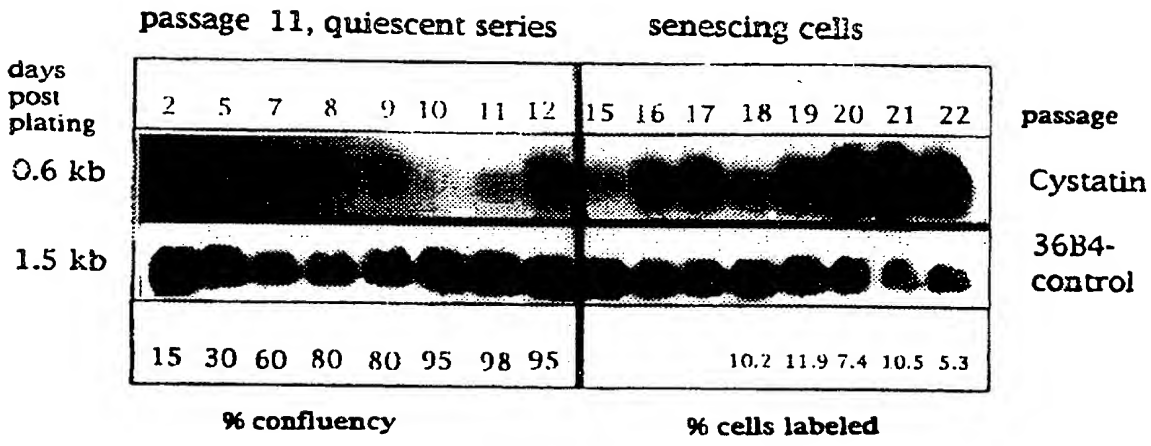

passage 11, quiescent series
senescing cells
days post plating
2 5 7 8 9 10 11 12
15 16 17 18 19 20 21 22
passage
0.6 kb
Cystatin
1.5 kb
36B4-control
15 30 60 80 80 95 98 95
10.2 11.9 7.4 10.5 5.3
% confluency
% cells labeled

FIG.5B

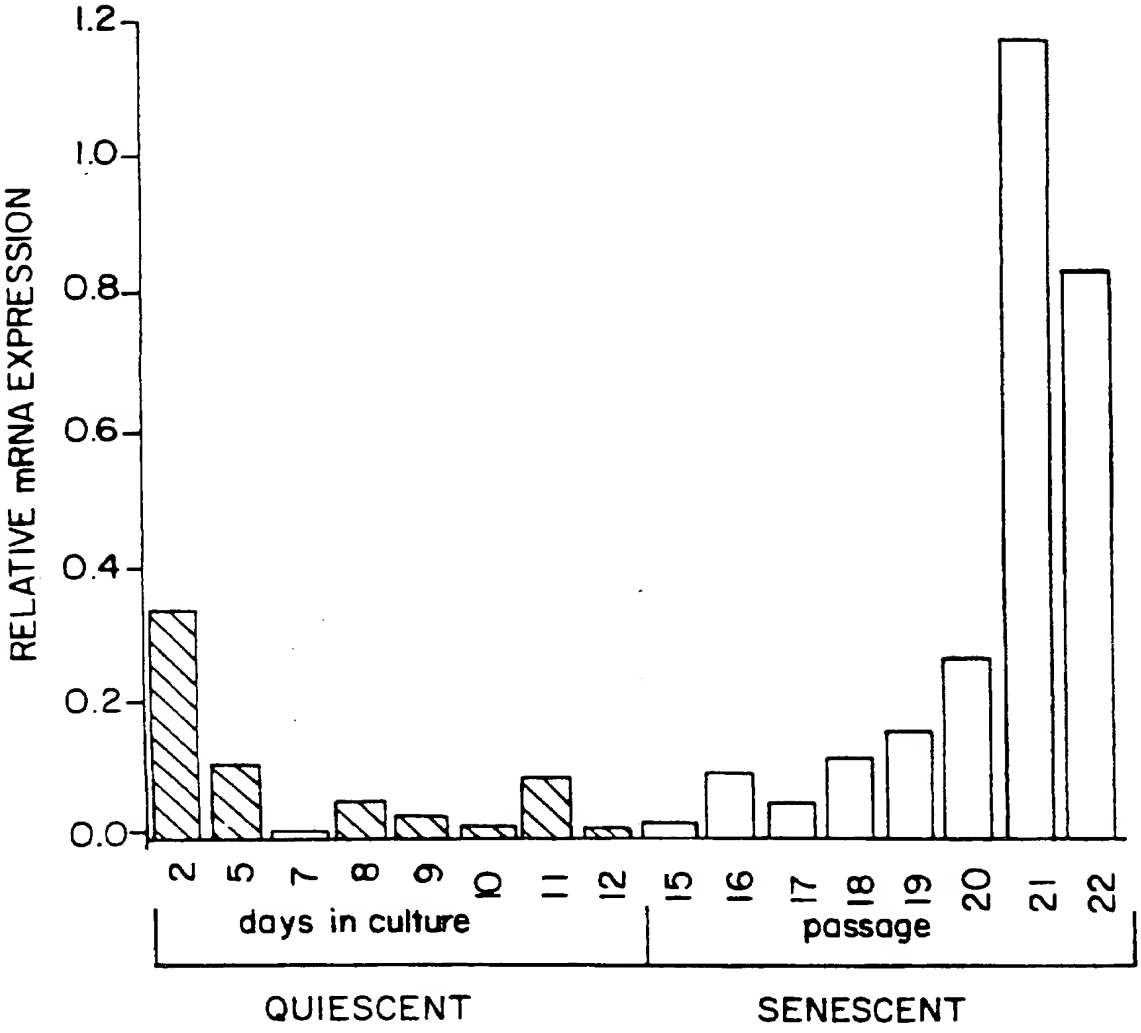

RELATIVE mRNA EXPRESSION
1.2
1.0
0.8
0.6
0.4
0.2
0.0
2
5
7
8
9
10
11
12
15
16
17
18
19
20
21
22
days in culture
passage
QUIESCENT
SENESCENT

FIG.6

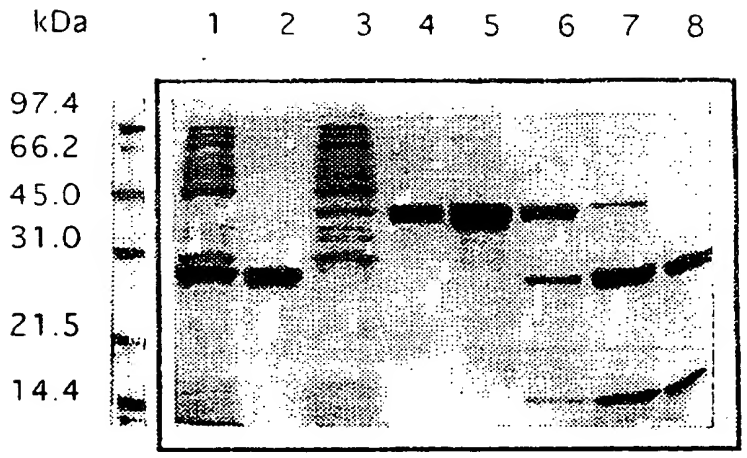


FIG.7A

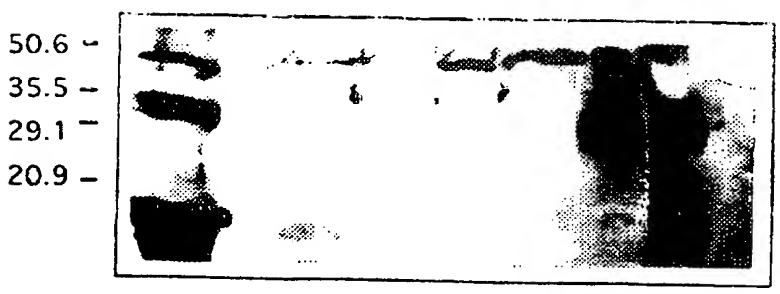

kDa
50.6
35.5
29.1
20.9
rGST-cys M
70N, L
21PT, L
MDA435, L
MDA157, L
BT549, L
70N, S
21PT, S
MDA435, S

FIG.7B

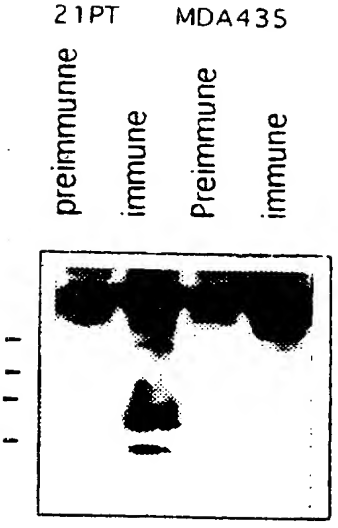

21PT
MDA435
preimmunne
immune
Preimmune
immune

FIG.8A

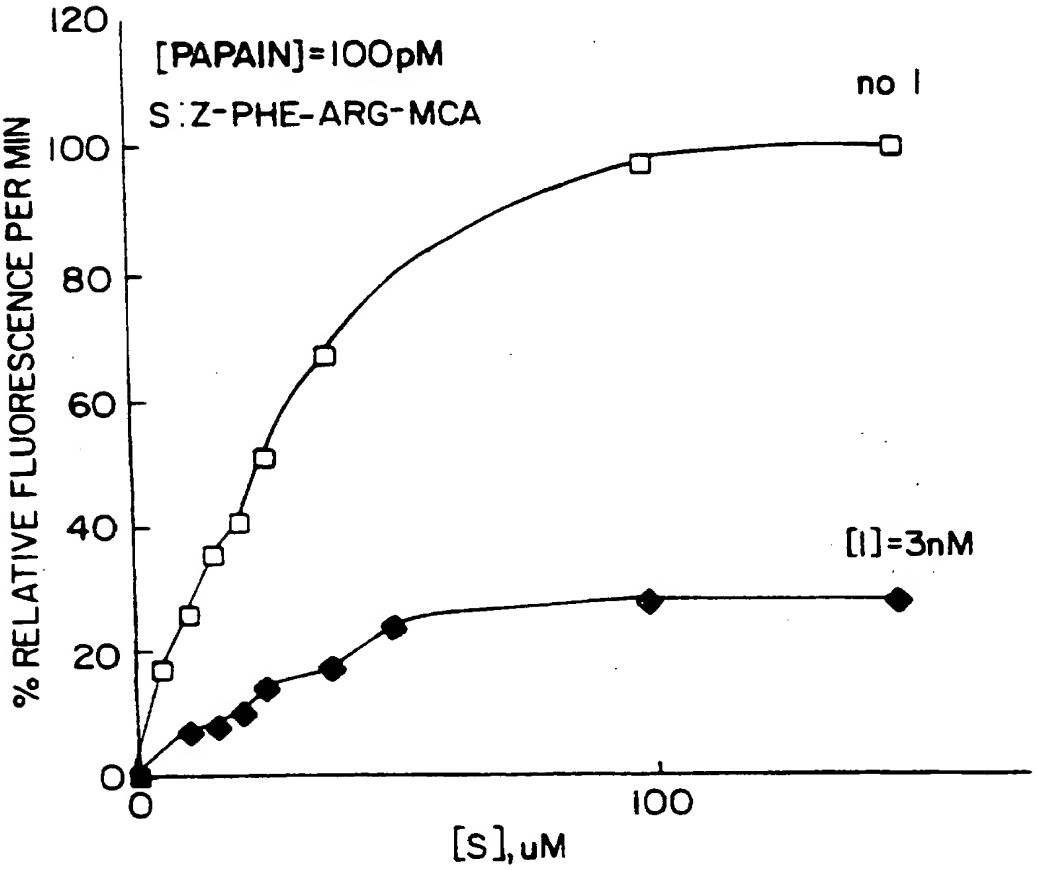

120
100
80
60
40
20
0
% RELATIVE FLUORESCENCE PER MIN
[PAPAIN]=100pM
S:Z-PHE-ARG-MCA
no I
[I]=3nM
0
100
[S], uM

FIG.8B

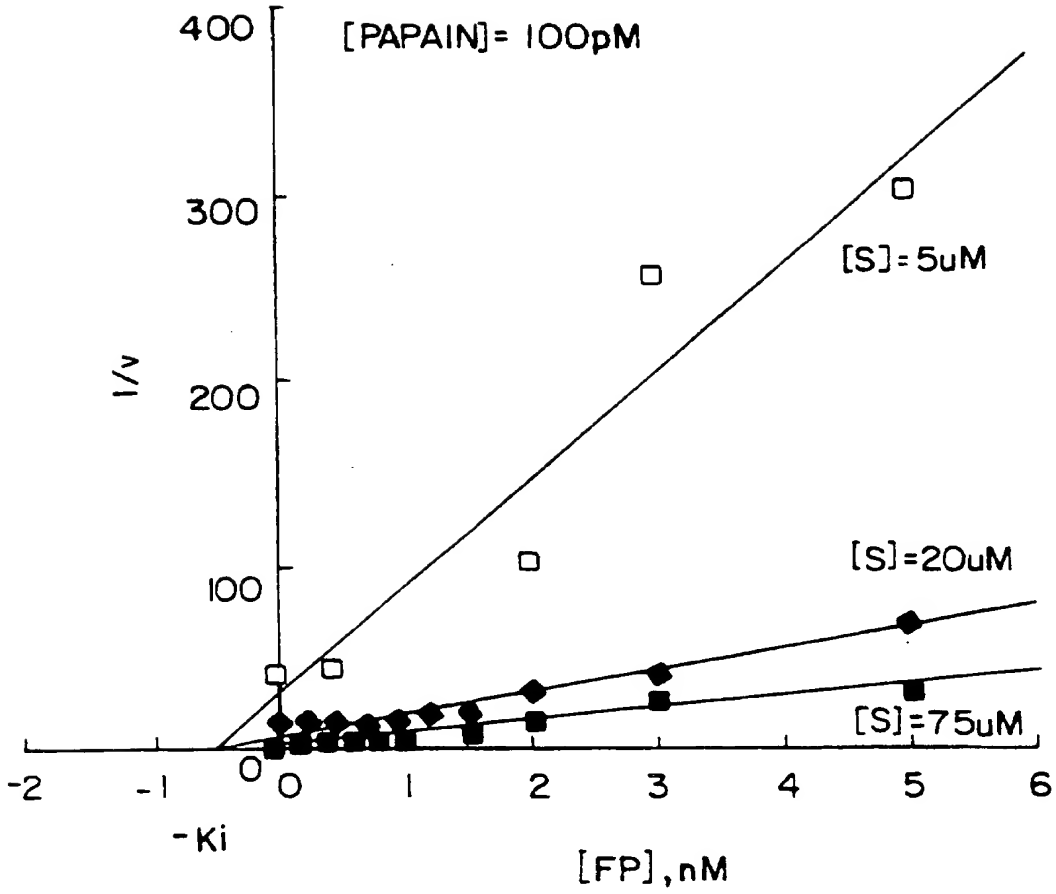
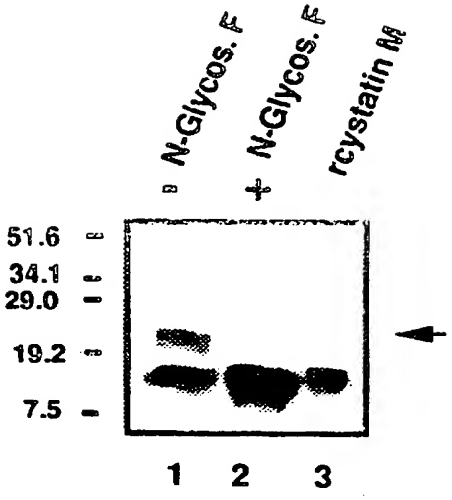

400
300
200
100
0
1/v
[PAPAIN]= 100pM
[S]=5uM
[S]=20uM
[S]=75uM
-2
-1
0
1
2
3
4
5
6
-Ki
[FP], nM

FIG.9



A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/15 C12N15/62 C12N15/11 C12N1/21 C12N5/10
C12N1/19 C07K14/81 A61K38/57 C07K19/00 C07K16/38
A61K39/395 A01K67/027 C12N15/00 G01N33/68 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE WPI Week 9528 Derwent Publications Ltd., London, GB; AN 95-212959 XP002031097 & JP 07 126 294 A (SNOW BRAND MILK PROD CO LTD) , 16 May 1995 see abstract see GENESEQ DATABASE, Accession no R78723 ---	1-4,6,7
A	DE 37 24 581 A (GRUENENTHAL GMBH) 2 February 1989 --- -/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 May 1997	12.06.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Van der Schaal, C

Form PCT/ISA/210 (second sheet) (July 1992)

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 A61K31/70 C12Q1/37

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X, O	1ST WORLD CONGRESS ON ADVANCES IN ONCOLOGY, ATHENS, GREECE, OCTOBER 22-26, 1995. INTERNATIONAL JOURNAL OF ONCOLOGY 7 (SUPPL.). 1995. 969. ISSN: 1019-6439, XP000672818 SOTIROPOULOU G ET AL: "Isolation of a novel protease inhibitor involved in breast cancer and senescence." see abstract ---	1-32
P,X	EMBL DATABASE, Accession no U62800. 26 Aug 1996. SOTIROPOULOU G. et al: "Identification, cloning and characterization of cystatin M ...." XP002031096 see the whole document -----	1-62

☐ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
	12.06.97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Authorized officer

Form PCT/ISA/210 (second sheet) (July 1992)

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 49-60
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claims 49-60 are partially directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
DE 3724581 A	02-02-89	EP 0330725 A	06-09-89
		JP 1124389 A	17-05-89